Identification of novel miRNAs as diagnostic molecules for detection of breast cancer using *in silico* approaches

UNIVERSITY *of the*
WESTERN CAPE

Najua Ali Ferrara

A thesis submitted in fulfilment of the requirements for the degree of Magister at the department of Biotechnology, Faculty of Science

University of the Western Cape

Supervisor: Dr Ashley Pretorius

October 2017

PLAGIARISM DECLARATION

I declare that *Identification of novel/miRNAs as diagnostic molecules for detection of breast cancer using in silico approaches* is my own work, that it 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 reference**.**

Signed:  . **Date :** 10/10/2017

N. Ferrara

ACKNOWLEDGMENTS

First and foremost, I praise Allah, the almighty for providing me this opportunity and granting me the capability to proceed successfully. You have given me the power to believe in myself and pursue my dreams and blessing in completing my studies. I could never have done this without the faith I have in you, the Almighty.

I would like to thank the valuable assistance I received from the following people. Their contributions towards the success of this project were perceived.

I would like to express my sincerest gratitude to my supervisor, **Dr Ashley Pretorius**, for his supervision, knowledge; guidance, encouragement expertise, understanding, and patience. Your belief in me and commitment to my success has been invaluable; and added considerably to my graduate experience. I acknowledge him for proofreading my thesis.

I would like to thank Dr. Mustafa Drah, Miss Chipampe Patricia Lombe for all the guidance and mentoring throughout my Master's degree. She is truly an incredible mind in bioinformatics. I greatly appreciate it.

I would like to thank all the students especially Miss Sharneal Lottering and Mr Riziki Martin.

I cherish the friendship I had and take this opportunity to thank them. Graduation friends have always been an encouragement and their motivation boosted my confidence that has helped me reach my goals. My heartfelt thanks to my close friends Mr.Naser Al Jahmi, Miss Hager Al Shaafi and Mrs Amane Allalgee for their support.

My sincerest gratitude goes to the Libyan government and DST/ Mintek Nanotechnology Innovation Centre NIC for the funding and financial aid. A grateful acknowledge.

I would also like to say a heartfelt thank you to my mother and father for always believing in me and encouraging me to follow my dreams. And my sister and my brothers for helping in whatever way they could during this challenging period.

DEDICATION

For my father and my mother

ABSTRACT

Breast cancer (BC) is the most common cancer in women worldwide, and is the second most common cancer in the world, responsible for more than 500 000 deaths annually. Estimates are that 1 in 8 women will develop BC in their lifetime. In South Africa, BC in women affects about 16.6 % of the population and could see a 78 % increase in cases by 2030. The failure of conventional diagnostic tools to detect BC from an early onset has revealed the need for diagnostic tools that would enable early diagnosis of BC. The current diagnostic tools include breast self-examination, mammography magnetic resonance imaging, ultrasonography and serum biomarkers; BRACA1, BRACA2, HER2. These conventional methods lack sensitivity, specificity and positive predictive value, and some of these diagnostic tools may be expensive and quite invasive. Therefore, novel diagnostic tools such as microRNAs which address the short comings of current methods are required for early diagnosis as well as BC management. MicroRNAs are a class of non-coding RNA molecules, which are important in RNA stability and gene expression. Various methodologies have been employed to identify novel microRNAs for diagnostics such as bioinformatics, also referred to as *in silico* analysis. The aim of this study is to identify novel microRNAs that can potentially detect BC at its earliest stage.

MicroRNAs (n=757) were extracted from various databases and all duplicates removed. These microRNA's were rigorously screened using several *in silico* methods identify those that have been previously validated as biomarker for BC thus leaving 27 microRNAs that could be potential novel biomarkers for BC management.

Target genes associated with these 27 microRNAs were identified using TargetScanHuman and miRDB. Duplicate target genes were removed and prioritized according to a prediction score of 90 % or higher.After which then gene lists obtained from the various databases with more than 900 gene targets were excluded and finally, duplications eliminated. A final list was generated containing 1,876 gene targets for 14 microRNAs. The 13 microRNAs not producing target genes based on these criteria were excluded. Functional annotation analysis generated 160 genes for the 14 microRNAs, which were further validated for their novelty as BC biomarkers. Through protein/protein interaction analysis and differential expression, nine genes were prioritized as BC biomarkers. These genes are regulated by five microRNAs (MiR1-MiR5) which may play key roles in the development of BC and its progression. Databases used for this part of the study included DAVID, STRING and Gene Expression Atlas.

The nine genes including the miRNAs were further analysed for their prognostic value in BC. Each database showed different microRNA target genes expression to have a poor prognostic outcome for the disease based on the p-values ($p \leq 0.05$). However, *ACTN4* was predicted as a good marker using SurvExpress and Kaplan-Meier Plotter. Combining the target genes seemed to provide good value for BC outcomes prognostic predictors when using the SurvExpress and GOBO databases.

Using different bioinformatics tools, viz. *in silico* approaches for expression analysis of microRNAs and their target genes, identified MiR1 and MiR2 as potential new

prognostic and diagnostic signatures as circulating microRNAs for BC. MiR5 can also act as a novel biomarker for BC but require further study.

KEYWORDS: Biomarkers, Bioinformatic, Breast cancer, Early Diagnosis**,** Expression profiling**,** *In silico***,** MiRNAs**,** Validation

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Full Name
%	Percentage
°C	Degree Celsius or degree centigrade
~	Approximately
∞	Infinity
≥	Greater than or equal to
A	Adenosine – Adenine
ACS	American Cancer Society
ACTN4	Alpha-actinin-4
API	Application programming interface
AP1G1	AP-1 complex subunit gamma-1
AIDS	Acquired immunodeficiency syndrome
ATCC	American type culture collection
BC	Breast cancer
BLCA	Bladder urothelial carcinoma
Bp	Base pair
BRCA	Breast invasive carcinoma
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
BSA	Bovine serum albumin
BSE	Breast self-examination
CANSA	Cancer association of South Africa
CA 15-3	Cancer antigen 15-3
cDNA	Complementary deoxyribonucleic acid
CD2AP	CD2-associated protein
CG	Co-expressed gene
CIS	Carcinoma *in situ*
cm2	Square centimeter

DAVID	Database for annotation,visualization and integrated discovery
DCIS	Ductal carcinoma in-situ
DM	Data mining
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DSBR	Double-strand break repair
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
ER/ ESR	Estrogen receptor
ES	Embryonic stem cells
ESCA	Esophageal carcinoma
FDA	Food and Drug Administration
FISH	Fluorescence *in situ* hybridization
G	Gravitational force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEA	Gene Expression Atlas
GEO	Gene Expression Omnibus
GLOBOCAN	Database of IARC which presents estimates of the burden of cancer incidence, mortality and prevalence worldwide
GO	Gene Ontology database
GOBO	Gene expression-based Outcome for Breast cancer Online
GOPC	Golgi-associated PDZ and coiled-coil motif-containing protein
GSA	Gene set analysis
HER-2/neu	Human Epidermal growth factor Receptor 2
HGCN	HUGO Gene Nomenclature committee
HIP1	Huntingtin Interacting Protein 1

HIV	Human immunodeficiency virus
HNSC	Head and neck squamous cell carcinoma
HRT	Hormonal replacement therapy
HPRD	Human Protein Reference Database
IARC	International Agency for Research on Cancer
IDC	Invasive ductal carcinoma
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
ITTACA	Integrated tumour transcriptome array and clinical data analysis
KDD	Knowledge Discovery in Databases
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KMPlot	Kaplan-Meier-plotter
KICH	Kidney Chromophobe
KIPAN	Pan-kidney cohort (KICH+KIRC+KIRP))
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LB	Luria broth
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
microRNA	Micro Ribonucleic acid
miRmine	Human miRNA Expression Database
mm	Millimeter
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NA	Nucleic Acid
NAP32G	acidic nuclear phosphoprotein 32 family member E
NCBI	National Centre for Biotechnology Information
NCI	National Cancer Institute

ng	Nanogram
NIH	National Institutes of Health
NHGRI	National Human Genome Research Institute
nm	Nanometer
PCR	Polymerase chain reaction
P53	Tumour protein P53
PR	Progesterone receptor
qRT-PCR	Quantitative real-time polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNase	Ribonuclease
RP	Reverse primer
RPM	Revolutions per minute
RT	Room temperature
RB1	Retinoblastoma 1
RBL1	Retinoblastoma-like 1
RBL2	Retinoblastoma-like 2
STAD	Stomach adenocarcinoma
STES	Stomach and esophageal carcinoma
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SYT9	Synaptotagmin-9
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TMED10	Transmembrane emp24 domain-containing protein 10
TNM	Staging primary tumour lymph node metastasis
Tris	Tris (hydroxymethyl) aminoethane
U	Unit
UCEC	Uterine corpus endometrial carcinoma

US	Ultrasound
VEGFA	Vascular endothelial growth factor-A
WHO	World Health Organization
3'	3 prime end
5'	5 prime end

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

1. Literature review

1.1 Overview of cancer

Cancer, a deadly disease is characterized by uncontrolled cell division leading to abnormal cell proliferation (Cancer Facts, 2012). According to the National Cancer Institute statistics in 2012, 14 million new cases and 8.2 million cancer-related deaths were recorded worldwide which is further expected to rise to 22 million in the next two decades (National Cancer Institute, 2016). In developing countries cancer is a major problem and the second common cause of death (Xu *et al*., 2008). In a recent study, Hashim and co-workers revealed that in 60 countries, the death rate due to cancer reached 3.7 million cases in 2010 (Hashim *et al*., 2016) whil 56 % of new cancer cases and 63 % of cancer deaths were reported in less developed countries. In developing countries the rate of new cancer cases is expected to increase from 56 % to 60 % by 2030 (Ferlay *et al*., 2010; Jemal *et al*., 2010). In the African population, cancer is an emerging public health problem with 715,000 new cancer cases by 2008 out of which 542,000 cases resulted in death (Stefan *et al*., 2013; Jemal *et al*., 2012).

In a country specific variability of occurrence of cancer types in 184 countries worldwide, breast cancer (BC) has emerged as the leading cancer type in 73 countries, including parts of Central and South America, Africa, and Asia (Figure 1.1). On the contrary, in 34 countries prostate cancer turned out to be the most

frequent type. These countries were characterized by a high life expectancy and prevalent testing of prostate surface antigens (Atun *et al*., 2015). Similarly, cervical cancer was also found to be the most common cancer type in 26 countries, mainly within sub-Saharan Africa and parts of Southern America. In-addition, while in 18 countries including parts of Eastern Europe, Western Asia, Northern Africa, and Eastern Asia (including China), lung cancer is reported to be the most frequent cancer type, colorectal cancer incidence was found to be highest in Europe and in Eastern Asia.

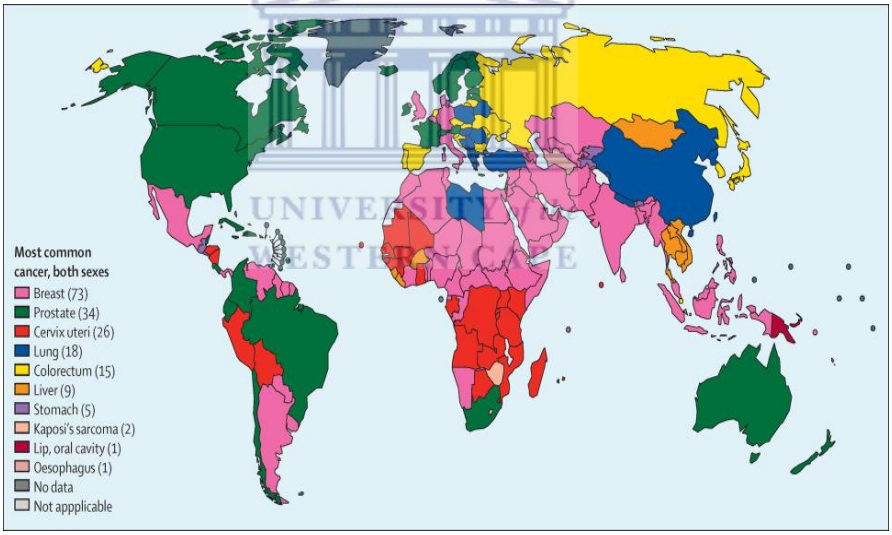


Figure 1.1: Estimated most frequent cancer types worldwide (Adapted from Atun *et al*., 2015).

1.2 Hallmarks of cancer

Tumours arise from normal cells through genetic alterations affecting the tightly controlled systems for growth control. More than a decade ago, Hanahan and Weinberg pointed out six hallmarks of cancer – (i) self-sufficiency in growth signals, (ii) insensitivity to growth inhibitory signals, (iii) evasion of apoptosis, (iv) limitless replicative potential, (v) sustained angiogenesis, (vi) tissue invasion and metastasis, all accepted universally in almost all human malignancies (Figure 1.2) (Pietras, and Arne, 2010; Cook *et al*., 2005). However, research advancements expand the understanding of these hallmark capabilities with new pathways discovered in cancer development (Gutschner & Diederichs, 2012).

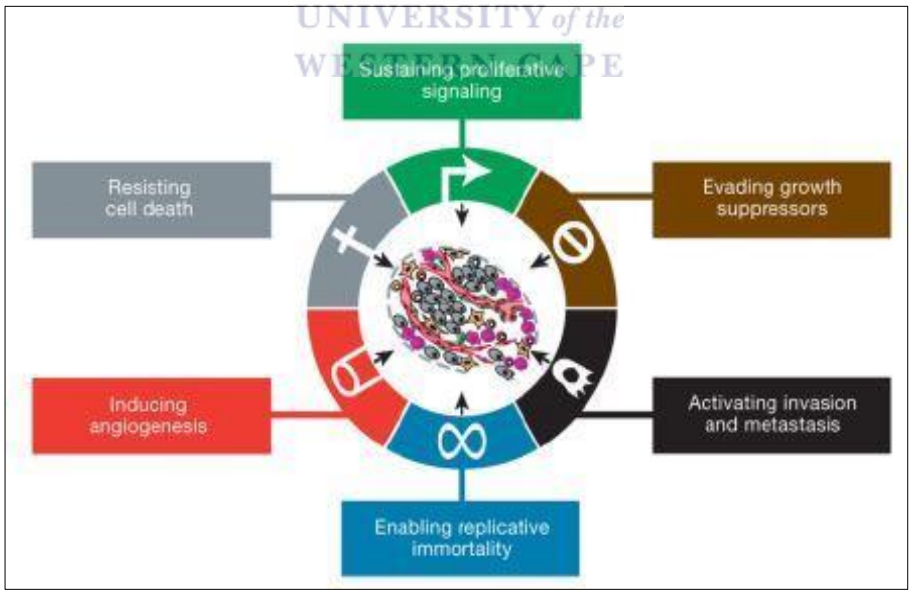


Figure 1.2: Six cancer hallmarks undertying transformation of normal cells to a neoplastic state (Hanahan & Weinberg, 2011).

1.2.1 Sustaining proliferative signalling

The most fundamental characteristic of cancer cells is their ability to sustain chronic proliferation. While normal tissues carefully control the production and release of growth-promoting signals that regulates the cell growth-and-division cycle, these signals get disturbed in cancer cells (Hanahan & Weinberg, 2011). In addition, tumour cells generate most of their own growth signals through the alteration of extracellular growth signals, trans cellular transducers of those signals, or intracellular circuits that translate these signals into action (Mees *et al*., 2009). These altered signals are transmitted by growth factors which in turn binds to intracellular tyrosine kinase domains of cell-surface receptors resulting in an aberrant cell cycle and growth. These signals often influences other cell-biological properties, such as cell survival and energy metabolism. Therefore, tumour cells depend to some extent on exogenous growth stimulation (Hanahan & Weinberg, 2011).

1.2.2 Evading growth suppressors

Cancer cells have the ability to evade anti-proliferative signals, such as soluble growth inhibitors and immobilized inhibitors, either by forcing cells out of the active proliferative cycle into the G0 phase of the cell cycle or by inducing cells to permanently give up their proliferative potential and forcing them to enter into post mitotic states (Lando,2011).

At the molecular level, most of these proliferative signals are funnelled through retinoblastoma 1 (RB1) and its two relatives, retinoblastoma-like 1 (RBL1) and retinoblastoma-like 2 (RBL2) proteins. Disruption of the RB1 pathway allows proliferation and thus renders cells insensitive to antigrowth factors, that normally operates along this pathway, to block cellular progression (Hanahan & Weinberg, 2000). Furthermore, cancer cells also avoid the powerful programs that negatively regulate cell proliferation and many of these programs depend on the actions of tumour suppressor genes. The two quintessential tumour suppressor genes encode the retinoblastoma (RB) and tumour protein P53 (P53) proteins which operate as central controls within two complementary cell regulatory circuits of cell proliferation (Chaffer & Weinberg, 2011; Hanahan & Weinberg, 2011). Hence, tumour cells with defects in the RB or P53 pathways lack important gatekeepers of cell cycle proliferation. This may cause cells to cease proliferation and enter the G 0 (rest) phase of the cell cycle (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011).

1.2.3 Resisting cell death

Apoptosis represents an essential mechanism to avoid the accumulation of damaged cells and mutations that are likely to culminate in the process of cancer formation (Fulda, 2009). In normal cells this form of programmed cell death is initiated by a variety of physiological signals that will unfold a sequential series of steps, including disruption of cellular membranes, breakdown of cytoplasmic nuclear skeletons, degradation of chromosomes, and fragmentation of the nucleus (Fulda, 2009). Cancer

cells are able to evade this induction of cell death perhaps due to the presence of survival signals supplied by the stromal compartment (Hanahan & Weinberg, 2011). Furthermore, cells use a variety of ways to avoid cell death, the most common one is mutations within the tumour suppressor p53 gene. The loss of p53 gene activity in cancer cells is paralleled by the increased synthesis of p53 inhibitors, or silencing its activators. The PI3 Kinase-AKT/PKB pathway is another pathway used by tumour cells to avoid programmed cell death. This pathway is concerned with anti-apoptotic survival signals. This survival signalling circuit is found to be up-regulated either by extracellular factors such as IGF-1 and IL-3, or by intracellular signals like RAS, thus leading to evasion of apoptosis (Wong, 2011).

1.2.4 Inducing angiogenesis

Tumours secrete growth factors that promote the formation of new blood vessels (angiogenesis) in order to support the growth of cancerous cells by supplying oxygen and nutrients. These growth factors stimulate proliferation of endothelial cells within the walls of capillaries in surrounding tissue with consequent production of new capillaries in tumours (Hanahan & Folkman., 1996). Furthermore, these actively growing new capillaries help the cancer cells to enter the circulatory system and begin the metastatic process (Garraway & Lander, 2013).

1.2.5 Activating invasion and metastasis

Invasive and metastatic cancer cells acquire the ability to migrate from the original tumour site to various organs. Changes promoting invasion and metastasis happen at the cellular level that include changes in the expression of surface markers which in turn allow the cells to adhere to the neighbouring tissues (Hanahan & Weinberg, 2011).

1.2.6 Enabling replicative immortality

Restricted life cycle of normal cells is mainly attributed to two barriers of proliferation, namely; *senescence*, the irreversible entrance into a non-proliferative but viable state, and *crisis*, which causes cell death. Cancer cells rarely emerge from crisis and this transition is called immortalization, which is essential for an infinite replicative potential to produce macroscopic tumours (Hanahan & Weinberg, 2011). Furthermore, telomeres, which protect the ends of chromosomes, are also implicated in restricted cell proliferation. In normal cells, telomere is reduced progressively with each cell cycle and eventually fails to protect the chromosomal DNA (Jafri *et al*., 2016). In addition, telomerase activity is generally absent in normal cells but it may be highly expressed in immortalized cancer cells (Killela *et al.,* 2013). The presence of telomerase activity is directly correlated with resistance of both senescence and apoptosis (Simon *et al*., 2000; Hanahan & Weinberg, 2011). Moreover, activation of proto oncogenes due to genetic alteration can also induce tumourigenesis (Mayr *et al*., 2009). While proto-oncogenes in mammalians regulate normal cell proliferation

and differentiation, these genes have been frequently activated in cancer cells either by mutation of the encoded protein (Mayr *et al*., 2009).

1.3 Cancer risk factors

In developing countries, cancer is an emerging public health challenge. This is partly attributed to the aging and growth of the population and increased spread of cancer risk factors (Braithwaite *et al*., 2012). In Africa, the increased prevalence of cancer risk factors is associated with economic transition as well as adoption of western lifestyles. According to the United Nation's Population estimates a 50 % and 90 % increase in emergence of new cancer cases in the elderly population in 2010 and 2030 respectively is anticipated (Jemal *et al*., 2012). The limited resources and other pressing public health problems, including communicable diseases such as acquired immunodeficiency syndrome (AIDS)/human immunodeficiency virus (HIV) infection, malaria, and tuberculosis have added a burden to public efforts to address cancer (Jemal *et al*., 2012).

Cancer can be caused by external factors or internal factors. External factors that increase the risk of cancer includes: tobacco, chemicals, radiation or infectious organisms. Internal factors, on the other hand, include genetics, chemicals, hormones, immune conditions and mutations that occur from metabolism (Stratton *et al*., 2009).

In Africa the most important cause of lung cancer is smoking. In South Africa, for instance, smoking is responsible for 61 % of male and 48 % female deaths from lung cancer (Sitas *et al*., 2008). Obesity is another main risk factor for developing chronic diseases, including malignancies and several types of tumours such as breast, prostate, and kidney cancer (Sanchez-lara *et al*., 2009).

1.4 Breast cancer

Breast cancer is the second most common malignant tumour after non melanoma skin cancer with significant mortality (American Society 2015, Board, 2017; Ferlay, *et al*., 2015). In 2012, 1.7 million women were diagnosed with BC worldwide resulting in 522,000 deaths. While the rate of incidence was higher (71.7/100,000) in more developed countries in comparison to less developed countries (29.3/100,000), the corresponding mortality rates were 17.1/100,000 and 11.8/100,000, respectively. That is, deaths caused by BC are nearly 17 % higher in the less developed countries (Figure 1.3) (Tao *et al*., 2015). In the USA, an estimated 231,840 and 246,660 new cases of invasive BC were diagnosed amongst women in 2015 and 2016 respectively (American Society 2016). Furthermore, an estimated 40,730 deaths occurred amongst BC patients in 2015 which is likely to be increased to 40,890 in 2016 (American Society, 2015, 2016). In Australia, around 2 % of BC cases occurs in young women between 20 and 34 years of age and 11 % incidence rate amongst women aged 35 and 44 years of age were observed (Hickey *et al*., 2009). In the African population, BC is

reported to be the most common type of cancer amongst women with an age-standardised incidence rate of 27 per 100 000 women, and a major cause of cancer mortality accounting for 16 % of cancer deaths amongst women (Moodley *et al*., 2016). It is estimated that amongst the 92,600 BC reported cases death occured in 50,000 cases (Jemal *et al*., 2012). According to the National Cancer Registry of SA, in 2010, BC was reported in 3,157 cases amongst Blacks, 816 cases amongst Coloureds, 1,824 cases amongst Whites and 340 cases amongst Asians (Cancer Association of South Africa, 2015). In addition, BC is the most common cancer in pregnant and postpartum women, occurring in about 1 in 3,000 pregnant women (Risal *et al*., 2008).

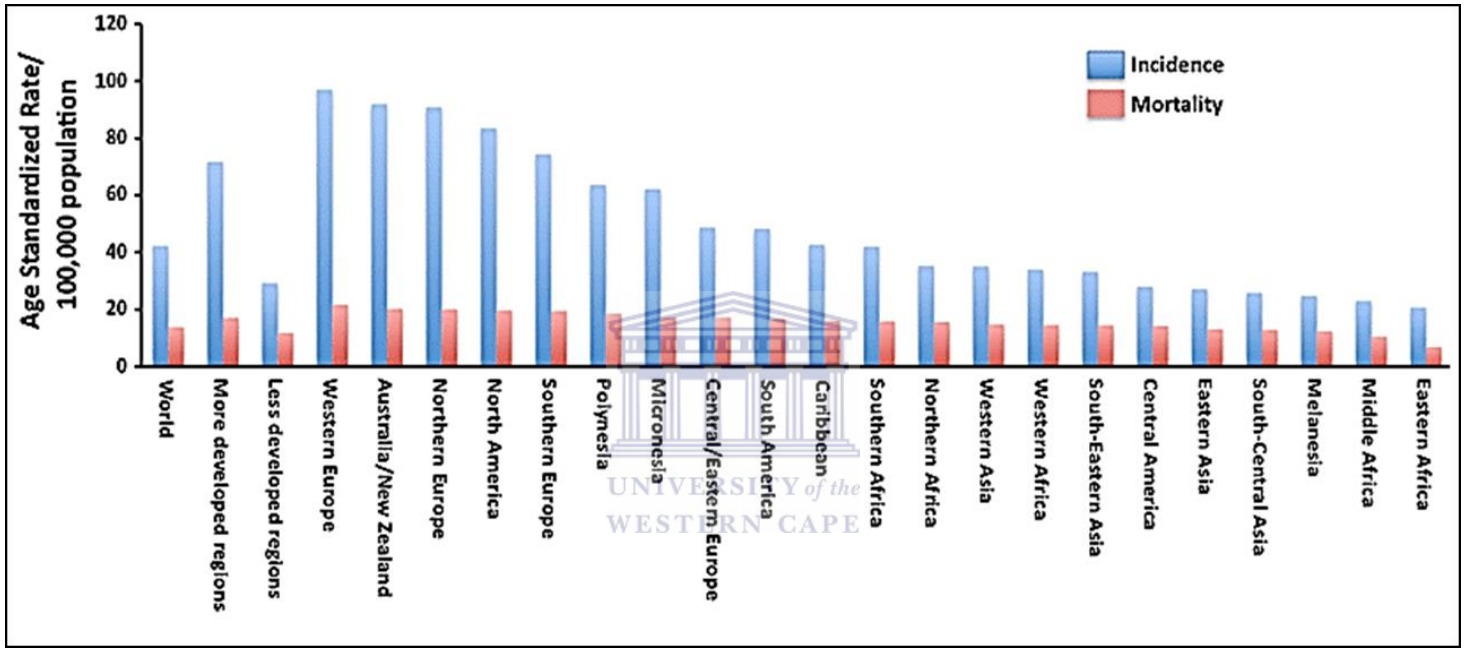


Figure 1.3: Estimated age-standardized rates based on 2008 GLOBOCAN data for incidence and mortality of female breast cancer by world region (Adapted from Tao *et al.*, 2015).

1.5 Types of breast cancer

Breast cancer can be classified into *in situ* and invasive (infiltrating) carcinoma based on the histological appearance of cancer cells. Breast carcinoma *in situ* (CIS) can either be ductal or lobular (Malhotra *et al*., 2010). Ductal carcinoma *in situ* (DCIS) are restricted within the duct and have not infiltrated the surrounding breast tissue (Bravaccini *et al*., 2013). During DCIS development, neoplastic cells grows uncontrollably into a lesion resulting in multiple layers which accumulate inside the ducts. The breast stroma consisting of the extracellular matrix, lymphatic's, blood vessels, stromal cells, immune cells and fat cells can either promote or suppress the carcinogenic process by responding to secretory signals (Espina *et al*., 2011). If not treated, DCIS may advance at a later stage to invasive ductal carcinoma (IDC), but the mechanism of this transition is not yet well understood (Sue *et al*., 2013).

Invasive ductal carcinoma (IDC) originates from the milk ducts and spreads through the walls of the ducts to surrounding breast tissues (Johnson *et al.*, 2012). It is the most common type of BC accounting for 72 to 80 % of all invasive BC (Arps *et al*., 2013) and 8 to 14 % of all BC cases (Malhotra *et al*., 2010). Invasive carcinoma is divided to six types namely, tubular, ductal lobular, infiltrating ductal, mucinous, medullary and invasive lobular carcinoma (ILC) (Malhotra *et al*., 2010). Invasive or infiltratory lobular carcinoma (ILC), is another major invasive tumour type that originates from lobules and compared to IDC, it is more probable to be positive to

hormone receptors (Arpino *et al*., 2004). ILC is characterized by a general thickening of the breast area, commonly the section above the nipple and toward the arm, which cannot be easily detected by mammography (Cao *et al*., 2012). If not treated within 3 years of disease diagnosis, the tumour cells will spread to different parts of the body such as the bones, lungs, liver and brain (Fernandez *et al*., 2013; Switzer *et al*., 2015).

1.6 Risk factors of breast cancer

Major risk factors of BC include genetic predisposition related to family history, and personal factors such as reproductive history and medical history (Tyrer *et al*., 2004). Individuals with a family history of BC are at a higher risk for disease development (Breast Cancer Now, 2017). Mutations within the tumour suppressor genes BRCA 1 (breast cancer type 1) and BRCA 2 (breast cancer type 2) are implicated in BC progression and are currently used as molecular markers for diagnostic and therapeutic intervention in BC patients (Petrucelli *et al*., 2013).

Women who have started menstruating early in life or who have late menopause were reported to be at an increasing risk of developing BC. An increased risk of BC may be associated with pregnancy at younger ages (before 30 years) While significant cancer progression is reported in *BRCA1* mutation carriers, breast feeding can reduce cancer risk in *BRCA1* rather than *BRCA2* mutation carriers (Friebel *et al*., 2014; Board, 2017; Kotsopoulos *et al*., 2012). Regarding the effect of pregnancy on BC,

neither diagnosis of BC during pregnancy nor pregnancy after BC seems to be associated with adverse survival outcomes in women who carry a *BRCA1* or *BRCA2* mutation (Valentini *et al*., 2013).

Hormonal history for women appears to be a risk factor, as the relative risk of BC seems to be related to the breast's cumulative exposure to estrogen and progesterone (Richie *et al*., 2003; Hulka *et al*., 2001). Furthermore, this connection between the risk factors of BC and the use of Hormonal Replacement Therapy (HRT) has been investigated for decades in many epidemiological studies (Collaborative Group on Hormonal Factors in Breast Cancer 1997). Many women who have used estrogen replacement therapy for extended periods of time (e.g. for more than 10 years) find a modest increase (approximately 3 % per year) in BC risk (Bernstein *et al*., 2002). Smoking and alcohol consumption are also inferred to increased the risk of BC (Gaudet *et al*., 2013).

1.7 Stages of breast cancer

Breast cancer staging systems are based on the clinical size and extent of invasion of the primary tumour (T), the clinical absence or presence of Palpable axillary lymph nodes and data of their local invasion (N), along with the clinical and imaging evidence of distant metastases (M). Afterwards, this staging is then translated into the TNM classification which has been further divided into five stages as follows:

Stage 0 called carcinoma *in situ* both lobular (LCIS) and ductal (DCIS). Ductal carcinoma *in situ* (DCIS) can be defined as a non-invasive condition in which abnormal cells are located in the lining of a breast duct. Usually, the cancerous cells do not spread from the duct to other breast tissues, however in some cases, DCIS may become invasive cancer. At this time, there is no way to know which lesions could become invasive (National Cancer Institute, 2016) (Figure 1.4).

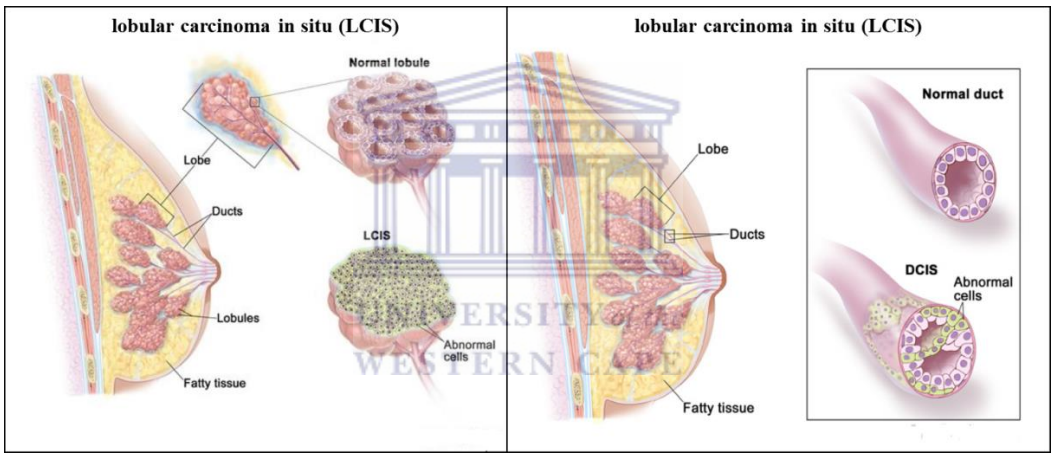


Figure 1.4: Stage 0, carcinoma *in situ* (lobular carcinoma in situ (LCIS) and ductal carcinoma *in situ* (DCIS) (Adapted from National Cancer Institute, 2016).

Stage I – This stage represent early stage BC where the tumour is less than 2 cm and is restricted to the breast. It is further divided into stages IA and IB, where stage IA represent a tumour of 2 centimetres or smaller and stage IB reflect small clusters of

BC cells (larger than 0.2 mm but not larger than 2 mm) in the lymph nodes (National Cancer Institute, 2016) (Figure 1.5).

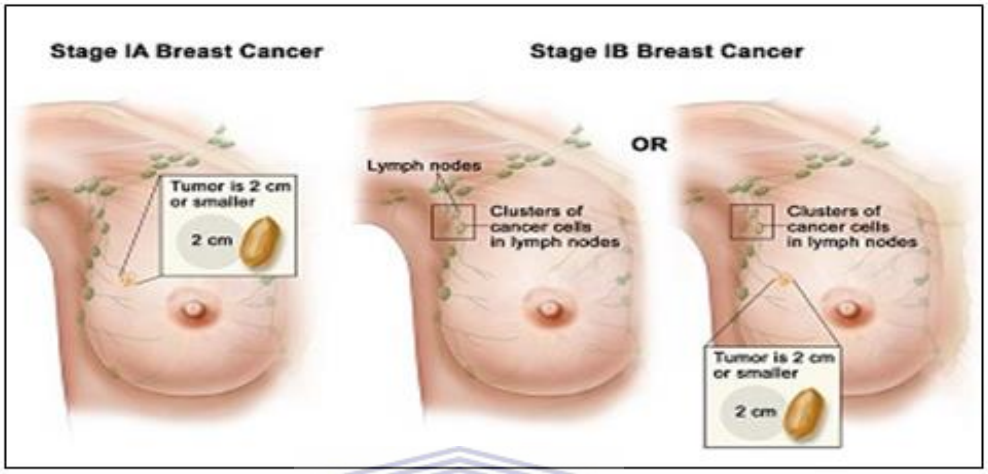


Figure 1.5: Stage I A and I B (Adapted from National Cancer Institute, 2016).

Stage II – Also known as early stage BC, stage II is characterized with a tumour size of 2 cm or less across. This stage tumour has spread to the lymph nodes under the arm; or the tumour of size between 2 and 5 cm (with or without spread to the lymph nodes under the arm); or the tumour is greater than 5 cm and hasn't spread to other organs. Like Stage I, Stage II is also divided into stages IIA and IIB (National Cancer Institute, 2016) (Figure 1.6). In stage IIA of BC a tumour is either absent in the breast or is less than or equivalent to 2 centimetres. Cancer (larger than 2 millimetres) is found in 1 to 3 axillary lymph nodes or in the lymph nodes near the breastbone (found during a sentinel lymph node biopsy); or the tumour size range within 2 – 5 centimetres. Cancer has not spread to the lymph nodes (National Cancer Institute,

2016). In stage IIB, the tumour size is between 2 – 5 centimetres. Small clusters of breast carcinoma cells (0.2 millimetres - 2 mm) are found in the lymph nodes with size ranging from 2 centimetres to 5 centimetres. (National Cancer Institute, 2016).

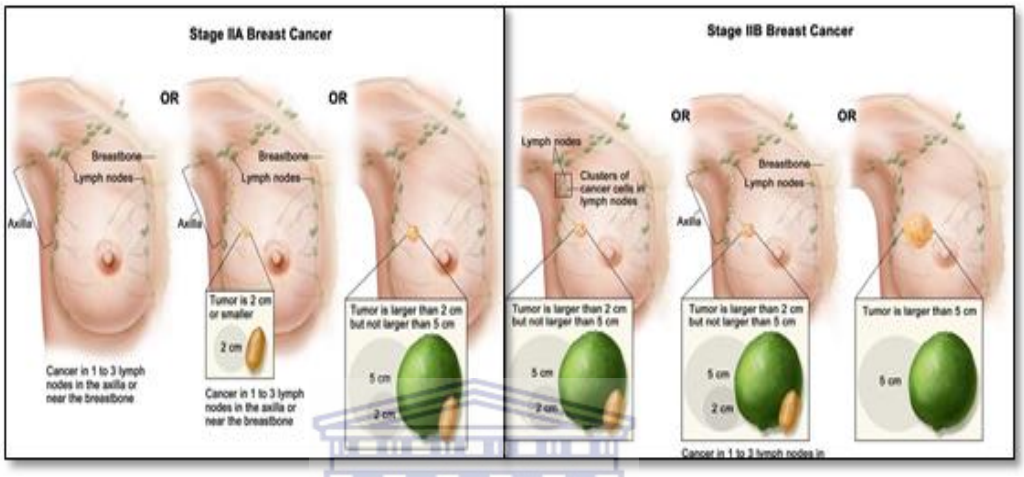


Figure 1.6: Stage II A and II B (Adapted from National Cancer Institute, 2016).

Stage III – Stage III illustrates locally advanced BC where the tumour is greater than 5 cm across and has spread extensively to the lymph nodes under the arm or near the breastbone or too the tissues near the breast (National Cancer Institute, 2016) (Figure 1.7). At stage IIIA a tumour is not present in the breast and cancerous cells has spread from 4 to 9 axillary lymph nodes or to the lymph nodes close to the breastbone (found during imaging tests or a physical exam). Small groups of BC cells (0.2 - 2 millimetres) are found in the lymph nodes; the tumour is sometimes larger than 5 centimetres. At stage IIIB, the size of tumour varies and cancer has spread to the chest wall with swelling or an ulcer on the skin of the breast. Also, cancer may have spread to up to 9 axillary lymph nodes; or the lymph nodes near the breastbone

(National Cancer Institute, 2016). At stage IIIC, a tumour is absent in the breast and cancer may have spread to the skin of the breast and caused swelling or an ulcer and/or has spread to the chest wall. Cancer are found to have spread to 10 or more axillary lymph nodes; or lymph nodes over or under the collar bone; or axillary lymph nodes and lymph nodes near the breastbone (National Cancer Institute, 2016).

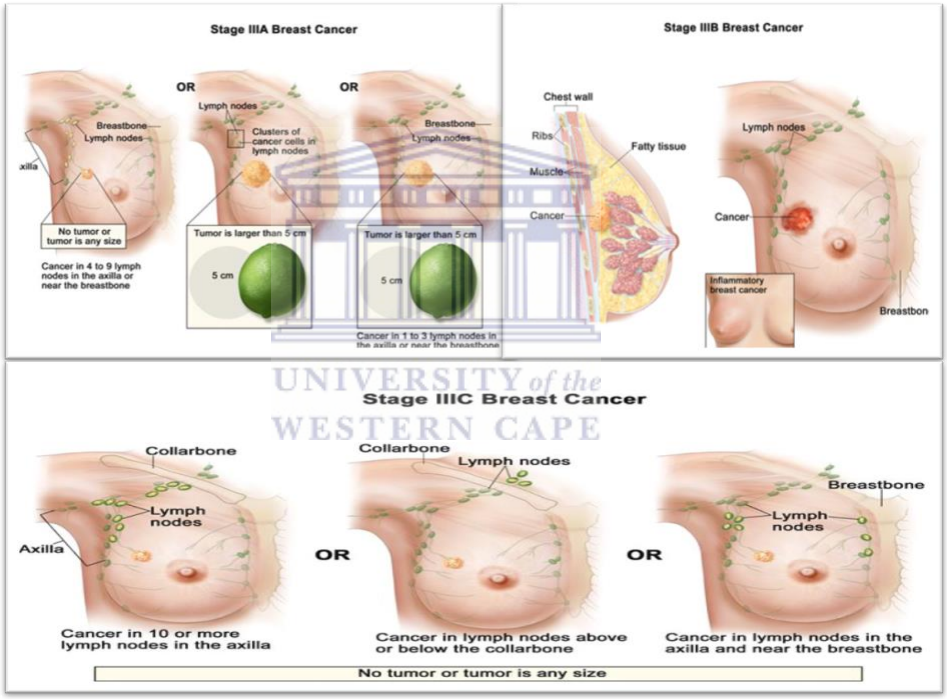


Figure 1.7: Stage III A, III B and IIIC (Adapted from National Cancer Institute, 2016).

Stage IV – Metastatic breast cancer, where the cancer has spread outside the breast to other organs in the body (National Cancer Institute, 2016) (Figure 1.8).

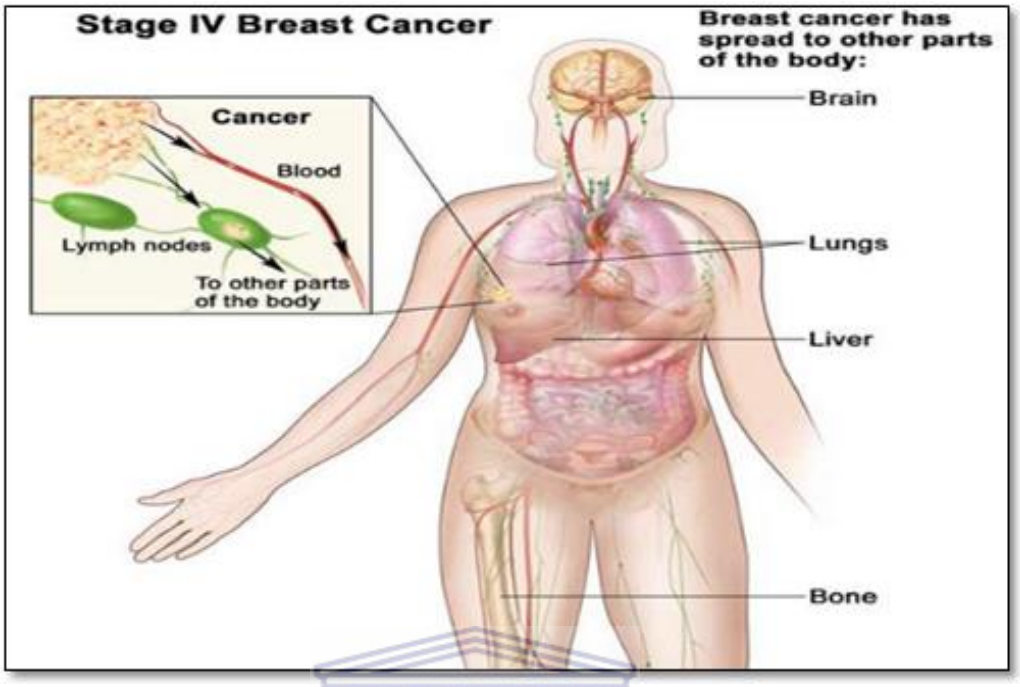


Figure 1.8: Stage IV (Adapted from Adapted from National Cancer Institute, 2016).

1.8 Diagnostic tools for breast cancer

The Following methods are extensively used in BC diagnosis:

1.8.1 Breast self–examination

This is the most preliminary test used by women to check for any irregularities in breast tissue (Petro-Nustas *et al*., 2013). The method involves physical examination of the breast by using finger pads in order to evaluate any possible lumps in the breast tissue as recommended by the American Cancer Society (ACS) (Evans, 2012).

However, There is no evidence on the effect of detection of BC during breast self-examination (BSE), because it is highly inaccurate.

1.8.2 Mammography

Mammography is a basic screening method for superficial diagnosis of BC especially for women aged between 39 and 69 years (Berg, 2009; Brodersen *et al*., 2010; Nelson *et al*., 2009). Although, mammography in combination with superior diagnostic methods has been reported to decrease mortality rates associated with BC, it has a limited impact due to poor sensitivity of the technique to detect BC (Evans, 2012; Mittal *et al*., 2017).

1.8.3 Ultrasound

Ultrasound (US) is more sensitive than a mammogram especially in patients with dense breast tissue (Madjar 2010; Luparia *et al*., 2013). In developing countries, US is becoming popular among lower level health centres (Gonzaga, 2010) due to its vast a vailability, inexpensive nature and non-invasive characteristics (Kelly *et al*., 2010). The limitations of Ultrasound are that most of the cancers cannot be detected. Ultrasound only detect an abnormality but it won't differentiate between cancers or benign conditions. Additional procedures are recquired for differential diagnosis (RadiologyInfo,. 2016).

1.8.4 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is commonly used to assess presence of complex lesions in women with a high risk of developing BC (Warner, 2011). As a screening tool for BC, it has shown a higher degree of sensitivity compared to mammography (Saslow *et al*., 2007). However, the technology renders certain limitations like presence of false-negative and false-positive results, human error and / or patient's characteristics, such as age or how they perceive things (Saslow *et al*., 2007).

1.8.5 Molecular diagnostic methods

Molecular biomarkers serve as a reliable indicator of a biological state, behaviour and function of the cells (Strimbu *et al*., 2010). The molecular diagnosis of cancer is based on the detection of molecular changes associated with disease and the identification of biomarkers associated with these molecular changes. There are different molecular diagnostic methods which can be used to evaluate the presence of such biomarkers in a patient sample. This includes techniques such as Quantitative Real Time Polymerase, Chain Reaction (qRT – PCR), enzyme–linked immunosorbent assay (ELISA), Fluorescent *in situ* hybridization (FISH) and Immunohistochemistry (IHC).

FISH is a cytogenetic technique used to detect the physical location of a specific gene or nucleic acid sequences (DNA or RNA) in intact chromosomes and to assess if multiple copies of disease specific genes or nucleic acid sequences (e.g, HER-2 gene

of BC) are present (Tanke *et al*., 2005). In a BC patient, this technique can differentiate between malignant and pigment lesions (Nijhawan *et al*., 2012).

IHC has paramount significance in the medical field, especially for pathological disease diagnosis (Liu *et al*., 2011). It is a standard technique used to detect the expression of disease specific proteins using an antibody (Mahmoud *et al*., 2011). IHC is commonly used to evaluate HER-2 status in BC patients (Garrison *et al*., 2013).

Quantitative Real Time Polymerase Chain Reaction (qRT – PCR) or reverse transcriptase assays have become the most commonly used method for characterising gene expression patterns in different sample populations. It is a technique that collects and generates data in real time with progressive PCR cycles (Derveaux *et al*., 2010). The reliability of qRT – PCR lies in its sensitivity and ability to detect a single copy of a specific transcript. Some limitations of this technique are non-specific amplification, variations in amplification efficiencies and heteroduplex formation (Pfaffl, 2012).

1.9 Breast cancer treatment

1.9.1 Surgery

Surgical removal of BC tissues usually comprises the first line of treatment followed by radiation, chemotherapy and hormonal therapy (Collins *et al*., 2011; Li *et al*., 2011). Many patients undergo a mastectomy, a surgical procedure which involves the removal of the whole breast (Coopey *et al*., 2013). In the case of early diagnosis, the removal of the cancerous tissue dissenting whole breast is preferred (Osman *et al*., 2013). The breast-conserving surgery pose side effects like pain or tenderness, impermanent swelling, formation of hard scar tissue at the surgical site, a change in the shape of the breast, a neuropathic pain arising in the chest wall, armpit, and/or arm that doesn't go away over time as with all operations, blood loss and infection at the surgery site (American society, 2016). Sometimes when axillary lymph nodes are also removed, side effects like lymphedema may occur (American society, 2016).

1.9.2 Radiation therapy

Radiation therapy after breast–conserving surgery can significantly reduce the risk of cancer recurrence as well as mortality among BC patients (Buchholz, 2009). The adverse effects of irradiation therapy include toxic effects of radiation depending upon the time and intensity of exposure and appearance of cardiovascular disease (Darby *et al*., 2013; Whelan *et al*., 2010). Furthermore, it is recommended not to use this therapy amongst patients suffering from systemic lupus erythematous,

scleroderma and pregnant women especially in the first trimester of their pregnancy (Buchholz, 2009; Whelan *et al*., 2010).

1.9.3 Chemotherapy

Chemotherapy involves the use of anti-cancer drugs for depleting the cancer cells (Smith *et al*., 2002). While chemotherapy is beneficial to premenopausal women suffering from BC, the treatment modality is associated with side effects such as vomiting, hair loss, and nausea which overall incur a negative impact on the patient's quality of life (Law *et al*., 2014; Dellapasqua *et al*., 2005). Although younger women have shown a higher preference for chemotherapy compared to older women, studies revealed better treatment outcomes for adjuvant therapy in women with BC (Niikura *et al*., 2013).

1.9.4 Endocrine therapy

Endocrine therapy (hormone therapy) decreases disease-related mortality and improves disease-free survival in patients with an early stage of BC (Romond *et al*., 2005). As suggested by (Niikura *et al*., 2013), endocrine therapies can serve as a better treatment strategy in comparison to chemotherapy. The authors also advised adjuvant endocrine therapy as a preferred treatment option in women with hormone receptor positive BC (Niikura *et al*., 2013). An example of this is tamoxifen, which reduces BC recurrence and mortality in hormone- positive BC patients (Connor *et al*., 2013).

1.9.5 Targeted therapy

Targeted therapy refers to drug or other substance molecules that specifically recognizes the cancer cells and blocks their growth and spread without harming the healthy cells (Nahta *et al*., 2006). Some of the targeting moieties used in the treatment of BC are monoclonal antibodies, tyrosine kinase inhibitors and poly (ADP-ribose) polymerase (PARP) inhibitors. Many targeted therapies approved by the Food and Drug Administration (FDA) include monoclonal antibodies such as trastuzumab (Herclon, Herceptin) and pertuzumab for BC patients (Mohamed *et al*., 2013).

1.9.6 Nanotechnology-based treatments

Nanotechnology has played a vital role in the treatment and diagnosis of cancer. Its significance lies in its ability to deliver chemotherapeutic agents directly to breast tissues. Nanoparticles ranging in size between 1 and 100 nm have a large surface area which makes them a highly efficient tool in targeted drug delivery systems (Yang *et al*., 2013). By targeting disease associated markers, it is possible to minimize negative effects to healthy tissue and increase drug efficacy at low dosages (Tanaka *et al*., 2009; Tharkar *et al*., 2015).

1.10 Cancer biomarkers

Biomarkers have a very important role in disease diagnosis, prognosis, predictive and treatment outcomes. While biomarkers help in differentiating physiological and

pathological mechanisms, they are equally important in assessing disease response to a medications therapeutic outcome, disease progression and to explore disease mechanisms (Strimbu *et al*., 2010). Furthermore, tumour biomarkers are substances in high concentrations in blood, urine or tumours. Such substances can be hormones, proteins, peptides etc. Tumour biomarkers can be specific or non-specific, making it useful in detection, diagnosis and prognosis of cancer (Dixon *et al*., 2000).

Diagnostic biomarkers help clinical oncologists in identification of risk factors and diagnose cancer at an early stage which is crucial for selection of the best treatment modality and monitor responses to treatment (Bhatt *et al*., 2010). Diagnostic biomarkers characterize abnormal biological processes and assess the patients who are at a high risk of disease development at a later stage (Madu and Lu, 2010; Tainsky, 2009).

Prognostic biomarkers can distinguish many stages of a disease and determine the course of therapy that should be applied to a particular patient after primary treatment (Riley *et al*., 2009). In BC, BRCA1, HER-2/neu, oestrogen receptor (ER) and progesterone receptor are few of the important prognostic biomarkers (Gerhardus *et al*., 2007; Mehta *et al*., 2012; Sapino *et al*., 2013).

Predictive biomarkers are used to determine relapse or the recurrence of disease after a patient had undergone treatment such as surgical removal of tumours (Mehta *et al*.,

2010). Oestrogen and progesterone receptors, HER2 etc. serve as important predictive biomarkers to envisage sensitivity to endocrine therapy and Herceptin treatment in BC respectively (Brünner *et al*., 2009).

1.11 Breast cancer biomarkers

Although several BC biomarkers have been developed their predictive value in diagnosis and detection of recurrence of the disease is a matter of debate (Alobaidi *et al.*, 2015).

1.11.1 Genetic markers

Breast cancer type 2 susceptibility protein (BRCA2) locus is the most studied genetic biomarker in BC (Wooster *et al.*, 1994). BRCA2 is a tumour suppressor gene and is involved in DNA repair pathways such as homologous recombination (Sharan *et al*., 1997; Stefansson *et al*., 2009). The expression of BRCA2 is observed in different cancer tissues such as prostate cancer (Castro *et al*., 2012). BRCA2 protein plays a significant role in the maintenance of genome integrity during Double-Strand Break Repair (DSBR) replication. This protein reacts with DNA repair proteins such as RAD51 Recombinase (RAD51) and Partner and Localizer of BRCA2 (PALB2). Failure to repair DNA damage can lead to replication errors resulting in DNA mutations and cancer (Schlacher *et al*., 2011). Pisano *et al*., (2011) have reported that

amongst the hundreds of mutations in the BRCA2 gene, several of its mutations have the potential to initiate BC.

1.11.2 Tissue markers

Human Epidermal Growth Factor Receptors (HER-2/neu also known as C-erB-2) discovered in 1985 was approved by the FDA in 1997 as a diagnostic biomarkers for BC using the FISH technique (Groudine *et al*., 1985; Ross *et al*., 2009). HER-2 is a proto-oncogene located on chromosome 17q21 which encodes for a 185 kDa protein and is a member of the tyrosine kinase family (Park *et al*., 2006; Rosian *et al*., 2005). It is reported to be over expressed in approximately 25-30 % of BC cases and its expression is particularly high in invasive BC specimens. Furthermore, over-expression of HER-2 has been associated with the negative expression of Estrogen Receptors (Mirtavoos-Mahyari *et al*., 2014). HER-2 is an important biomarker target for the treatment of the disease (Radojicic *et al*., 2011).

1.12.3 Serum markers

CA 15-3 is one of the most extensively studied circulating prognostic factors for BC. Preoperative concentrations in combination with existing prognostic factors are useful in predicting disease outcomes in patients with newly diagnosed BC (Alobaidi *et al*., 2015). CA 27- 29, however, is a less widely used serum marker in BC (Duffy & Michael, 2006). Proteins in the MUC-1 family which includes CA15.3, BR 27.29, MCA and CA549 are other commonly used serum markers. As these proteins possess

similar diagnostic sensitivities and specificities, the use of more than one MUC-1 antigen is not likely to be advantageous (Donepudi *et al*., 2014).

1.11.4 MicroRNAs

MicroRNA constitute a class of small noncoding RNAs that control gene expression by targeting mRNAs and triggering either translation repression or RNA degradation (Bartel *et al*., 2009). In particular, comparison of normal and malignant breast tissue has revealed that small subsets of deregulated microRNAs (including mir-125b, mir-145, mir-21, and mir-155) can unequivocally distinguish normal form malignant breast tissue. These differentially expressed microRNAs were found to be correlated with histopathological features of BC such as tumour size, nodal involvement, proliferative capacity and vascular invasiveness thereby enhancing the possibility that microRNA profiles may represent a promising new class of cancer biomarkers (Mattie *et al*., 2006). Furthermore, microRNAs are also emerging as new diagnostic biomarkers (Tijsen *et al*., 2012). In-extension, recent studies have confirmed that microRNAs have a unique expression profiles in different cancer types at different stages and they play an important role in many diseases and viral infections. These results suggest that microRNAs can function as novel biomarkers for disease diagnostics and constitute a new strategy for microRNA based gene therapy (Zhang *et al*., 2008).

MicroRNA genes are evolutionally conserved. They are either located within the introns or exons of protein-coding genes (about 70 %) or in the intergenic regions (30 %). While the expression of the intergenic microRNAs is related to their host gene expression, all intragenic microRNAs have independent transcription units suggesting that they are transcribed in parallel with their host transcripts (Acunzo *et al*., 2015). The second group of microRNAs is transcribed from intergenic regions or gene deserts which comprise independent transcription units. MicroRNAs are preferentially transcribed by polymerase II into long primary transcripts, up to several kilobases (pri-microRNA) that are subsequently processed in the nucleus by the enzyme Drosha to become ~70-nt-long precursor strands (pre-microRNA) (Figure 1.9) (DiLeva *et al*., 2014). At this point the Exportin 5 (XPO5) and its catalytic partner Ran-GTP bind the pre-microRNAs, and induce the export of this molecule from the nucleus into the cytosol and the pre-microRNAs is recognized in the cytosol by the RNAse III Dicer enzyme which processes the pre-microRNAs, producing a mature microRNAs duplex of about 22 nucleotides. At this step, the microRNAs double strand binds to the RNA-induced silencing complex (RISC) (Acunzo *et al*., 2015). RISC consists of the transactivation-responsive RNA-binding protein (TRBP) and Argonaute2 (Ago2). While it retains the mature strand fragment, the complementary strand on the other hand is removed and degraded, resulting in a fully functioning microRNA. MicroRNAs induce the degradation of their mRNA target through an imperfect complementary matching with the targets 3'UTR. Seed region of the microRNA is very important for their function and target specificity (Acunzo *et al*., 2015). In brief, the process of mature microRNA biogenesis is characterized by

three fundamental steps: cropping of pre-microRNAs, export and dicing, (Acunzo *et al*., 2015).

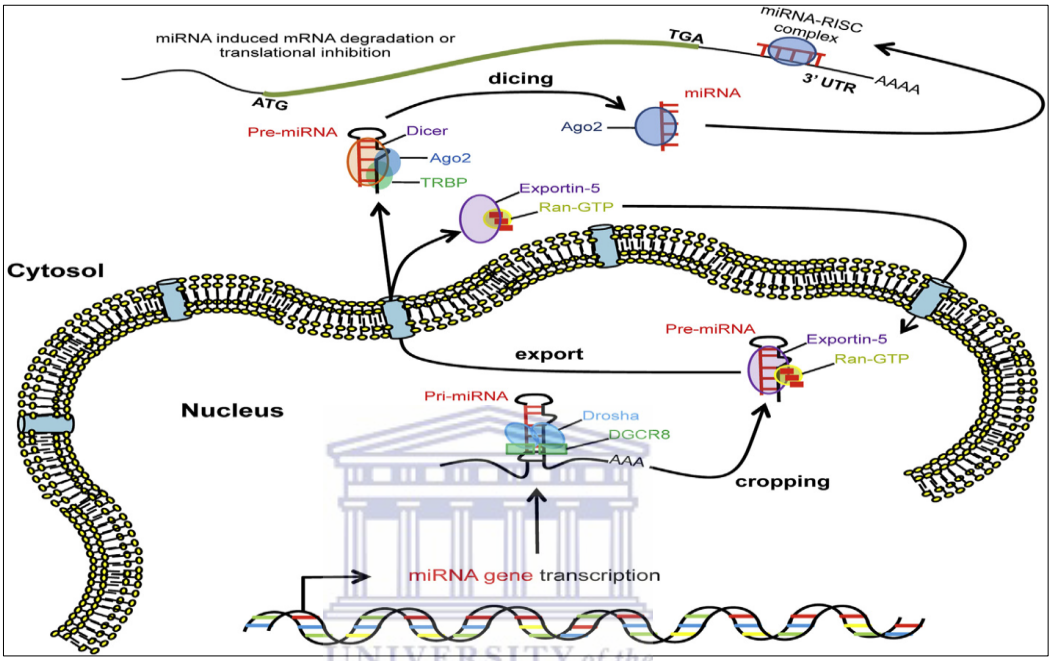


Figure 1.9: Schematic representation of microRNA biogenesis and miR-RISC activity (Adapted from Acunzo *et al*., 2015).

1.12 MicroRNAs as oncogenes and tumour suppressor genes

Alterations in oncogenes and tumour-suppressor genes have long been evident as a leading cause of tumourigenesis. It is indicated by studies that microRNAs can also serve as oncogenes or tumour suppressor genes in human cancer (Calin *et al*., 2004). While the expression of oncogenic microRNAs, popularly known as "oncomirs" is

up-regulated in tumours, the expression of tumour suppressive microRNAs (Ts-microRNA) is down-regulated (Zhang *et al*., 2007). The influence of oncomiRs and Ts-microRNA on cancer cell hallmarks has been reviewed (Grammatikakis *et al*., 2013).

1.12.1 MicroRNAs in cell growth and proliferation

Aberrations in the cell cycle regulators including cyclins and cyclin dependant kinases (CDKs) has been determined as one of the universal cause of cancer progression (Malumbres & Barbacid, 2001; Nigg, 1995). Oncogenic microRNAs target CDK inhibitors or retinoblastoma family proteins transcriptional repressors for increased cell proliferation and growth in cancer. Several microRNAs has been reported to be up-regulated, including human miR-17-92 cluster, miR-132, miR0212 in many cancers suggesting their role in undifferentiated proliferation (Lu *et al*., 2007; Park *et al*., 2011) (Figure 1.10). While considerable amount of evidence revealed oncogenic microRNA targets during cell cycle entry and G1/S transition, it may also affect the late phase of the mammalian cell cycle such as the G2/M phase (Chen *et al.*, 2010). In a study conducted by Bhattacharya *et al*., 2013 expression of Wee-1 was reported to be negatively correlated with miR-195 in melanoma patients and is responsible for unrestricted growth of tumour cells (Bhattacharya *et al*., 2013). It is evident that Wee1 is a negative regulator of the G2/M checkpoint that acts by cyclin B/CDK1 phosphorylation (Chen *et al*., 2010). In-addition, miR-128a, miR-155, and miR-516a-3p were also reported to regulate Wee1 expression in specific

cancers (Butz *et al.*, 2010). Similarly, miR-372, p53/miR-520g/p21, miR-520g, miR-106b family and miR195 were reported to regulate p21 protein expression which arrests cell cycle progression by inhibiting the activity of cyclin/cdk2 complexes (Qi *et al.*,2009; Gartel & Radhakrishnan, 2005; Zhang *et al*., 2015; Ivanovska *et al*., 2008).

Several microRNAs have been implicated that exert their oncogenic effect through the regulation of p53 expression (Hu *et al.*, 2010; Kumar *et al*., 2011; Le *et al*., 2009; Tian *et al*., 2010; Xu *et al*., 2010). It is evident that p53 protein regulates multiple genes involved in cell cycle arrest, apoptosis, DNA repair and senescence (Harris & Levine, 2005). While miR-504, miR-214 and miR-125b was reported to stimulate tumourigenesis by inhibiting p53 protein levels, miR-25 and miR-30d adversely affect the cell cycle process by targeting the 3'UTR of TP53 in several cancers (Hu *et al*., 2010; Kumar *et al*., 2011; Le *et al*., 2009; Xu *et al*., 2010). Moreover, mutations in p53 gene not only results in the loss of wild-type p53 tumour suppressor activity, but also modulate the expression of microRNA genes (Donzelli *et al*., 2012). In a study conducted by Ganci and colleagues, 12 microRNAs were identified which were involved with p53 mutations as well as cancer-specific survival (Ganci *et al*., 2 013).

Another mechanism through which several microRNAs have been implicated in promoting tumourigenesis is through inhibition of proto-oncogenes *MYC* and *RAS* (Schulte *et al*., 2008; O'Donnell *et al*., 2005; Ma *et al*., 2010; Mestdagh *et al*., 2010). Deregulation of these proto-oncogenes has been implicated in human malignancy (Cole & McMahon, 1999). MiR-92, miR-106a, let-7b, miR-17-5p, miR-93, miR-99

and miR-221 have been reported to influence *MYC* family expression in different cancer cells (Schulte *et al*., 2008; Mestdagh *et al*., 2010). Similarly, miR-155 and miR-21 were noted for oncogenic Ras-induced cell proliferation (Wang *et al*., 2015; Melnik, 2015).

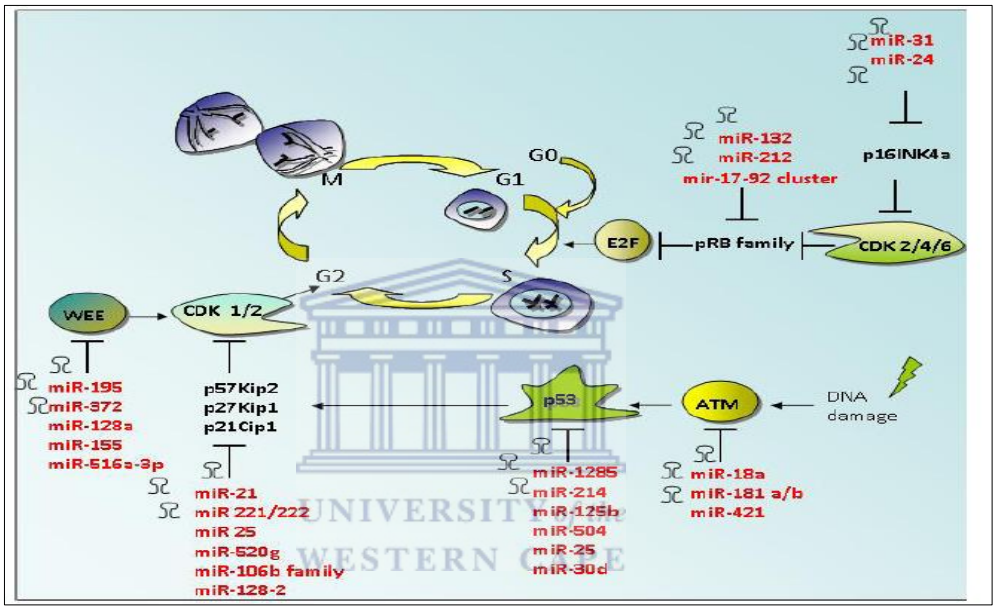


Figure 1.10: Oncogenic microRNAs involved in cell cycle progression (Adapted from Frixa e*t al*., 2015).

Aforementioned sections revealed that over-expression of oncogenic microRNAs result in enhanced tumourigenic effect mainly through stimulation of cell cycle check points causing shorter division periods (Witsch *et al*., 2010). In–addition, cancer cell growth and proliferation is also influenced by increased expression of proliferative signals such as B cell CLL/lymphoma 2 (Bcl-2), B-cell lymphoma-extra-large (Bcl-

xL), sirtuin 1 (SirT1), high-motility group AT-hook gene 1 (*HMGA1*), epidermal growth factor (EGF), vascular-endothelial growth factor (VEGF) and platelet-derived growth factor (PDGF) (Witsch *et al*., 2010). Many of these proliferative enhancers are regulated by Ts-microRNAs and down-regulation of these (tumor-suppressor (Ts-microRNAs)) results in uncontrolled growth and metastasis. MiR-34a, miR-125b and miR-17-5p were reported to regulate Bcl-2 and SirT1 expression responsible for cell proliferation and survival in cancer cells (Li *et al*., 2012; Sun *et al*., 2012; Yang *et al*., 1997). Another microRNA target essential for cell proliferation is RNA-binding protein HuR, which is reported to be influenced by miR-125a and miR-519 (Abdelmohsen and Gorospe, 2010; Abdelmohsen *et al*., 2010; Zhao *et al*., 2012). Suppression of miR-1, miR-28, miR-205, miR-296 has been implicated in multiple cancers. These microRNAs were reported to exert tumourigenic effect by enhanced expression of oncogenes and receptor tyrosine kinase MET and serine/threonine-protein kinase Pim-1 as well as reduced expression of tumour suppressor interleukins (Grammatikakis *et al*., 2013). Other potential TS-microRNAs involved in cell proliferation and growth in different cancers include miR-135a, miR-145, miR-128, miR-101, miR-143, miR-24, miR-133a, miR-133b, miR-138,miR-216b, miR-155, miR-138, miR-508-3p and miR-509-3p (Grammatikakis *et al*., 2013).

1.12.2 MicroRNAs in apoptosis

Alterations in apoptosis are one of the major hallmarks of cancer and are responsible for tumour initiation, progression and treatment resistance (Lowe & Lin, 2000).

Several studies indicated the role of oncogenic microRNAs in apoptosis and these microRNAs function by directly targeting pro-apoptotic mRNAs in both extrinsic and intrinsic pathways (Lima *et al*., 2011). Oncogenic microRNAs that have been implicated in regulation of the extrinsic pathway include miR-130a,miR-221, miR-222, miR-21, miR-18a, miR-144, miR-32 and miR-216a/217. Down-regulation of these oncogenic microRNAs results in tumourigenesis primarily through up-regulation of p27kip1, PTEN and TIMP3 proteins (Frixa *et al*., 2015). Similarly oncogenic microRNAs influencing intrinsic apoptotic pathway include miR-886-5p, miR-125b, miR-221 and miR-222, miR-181a, miR-17-5p-92, miR-106b-25, and miR-32. These oncogenic microRNAs were reported to be up-regulated in several cancers and reduce apoptosis by suppressing expression of various proliferative enhancers such as BAX, BAK1, BMF, PUMA and BIM proteins (Frixa *et al*., 2015). In addition to extrinsic and intrinsic pathways, microRNAs were also reported to regulate apoptosis through caspase expression (Floyd *et al*., 2014; Hudson *et al*., 2013; Tsang *et al*., 2008). In particular, miR-106b-25 cluster have been indicated to be associated with disease recurrence through caspase 7 inhibition (Hudson *et al*., 2013). Recently, other microRNAs such as miR-582-5p and miR-363, miR let-7a were also implicated as key apoptotic players in several carcinomas by directly targeting Caspase-3, Caspase-9, and BIM proteins (Floyd *et al*., 2014; Hudson *et al*., 2013; Tsang *et al*., 2008). (See Figure 1.11).

With regard to Ts-microRNAs, their down regulation results in the suppression of anti-apoptotic factors like Bcl-2, and p53 in cancer cells (Sigal & Rotter, 2000; Lukashchuk & Vousden, 2007). Decreased expression of miR-34a was reported to

reduce p21 and Bax expression through p53 with a concurrent increase in expression of SirT1, Bcl-2 or cyclin D1 (Zenz *et al*., 2009). Other targets of reduced miR-34a expression that decrease apoptosis and enhance tumourigenesis in cancer cells include baculoviral IAP repeat-containing 3 (BIRC3), decoy receptor 3 (DcR3), pro-survival factors c-Met, Notch-1 and Notch-2 (Guessous *et al*., 2010), Cyclin D1, Cyclin E2, Cdk4, Cdk6 and E2F (Liston *et al.*, 2003; Chen & Hu, 2012; Tazawa *et al.*, 2007). Down regulation of several additional microRNAs such as miR-125-b, miR-181d, miR-451 and miR-519 were also reported to suppress apoptosis and increased cell survival through aberrant modulation of Bcl-2 expression, Cyclin D1, p53, p21 and DNA repair enzymes (Wang *et al*., 2012; Nan *et al*., 2010; Abdelmohsen *et al*., 2012; Marasa *et al*., 2010).

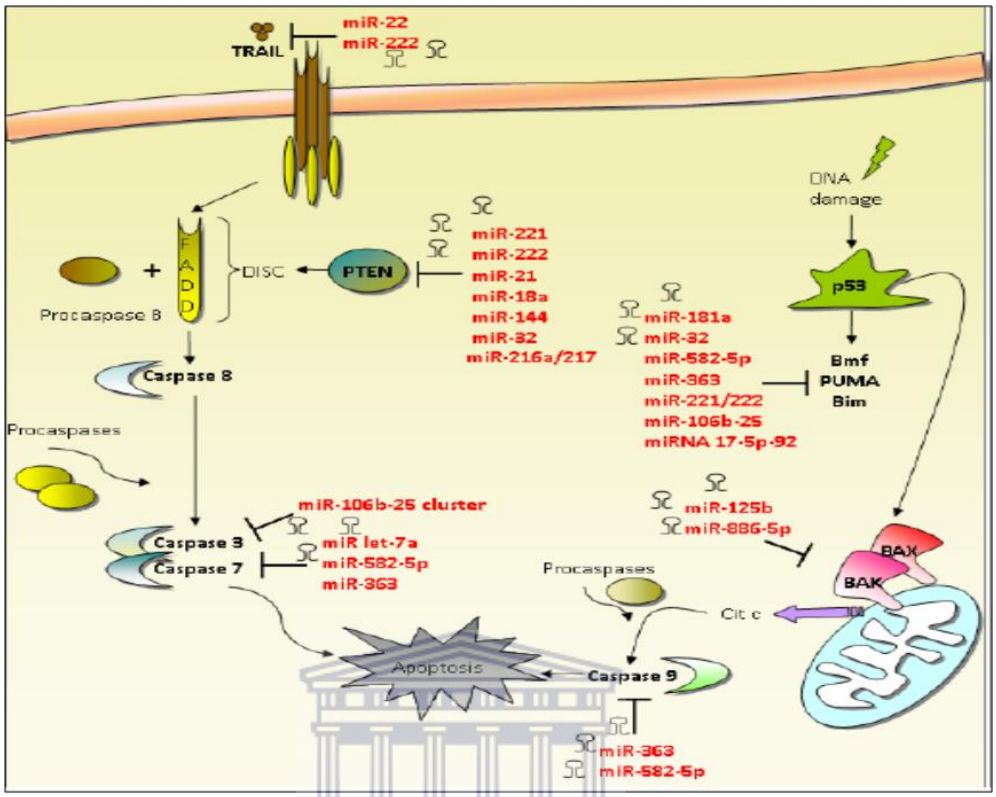


Figure 1.11: Oncogenic microRNAs involved in apoptotic pathways (Adapted from Frixa *et al.*, 2015).

1.12.3 MicroRNAs in metastasis

Oncogenic microRNAs involved in pro and anti-metastatic effects have been termed as metastamirs (Hurst *et al.*, 2009). Several microRNAs such as miR-21, miR-141, miR-10b, miR-9, miR-373, miR-885-5p and miR-520c were reported to be over expressed in different cancer types by targeting several metastasis suppressor genes such as TPM1, PDCD4, E-cadherin and Maspin as well as other genes like HOXD10,

and adhesion molecule CD44, thereby promoting invasion and metastasis (Frixa *et al*., 2015).

With regards to Ts-microRNAs, their suppression in cancer cells facilitate the expression of many pro-invasion and pro-metastatic proteins like E-cadherin repressors - Snail, Twist and Zinc finger E-box binding homeobox (ZEB1/2), MMPs, c-Met and VEGF, HMGA1, Bcl-2, SirT1, N-Ras, K-RAS, Ezrin, Mucin 4, E2F and metastasis-associated gene (MTA1) (Grammatikakis *et al*., 2013). Down regulation of miR-9, miR-200c, miR-340 and miR-34a, miR-29b, miR-205 in several cancers were reported to promote metastasization through aberrant regulation of anti-metastatic proteins E-cadherins and SNAIL, ZEB1 and ZEB2, c-Met, MMP-2, and VEGF-A respectively (Grammatikakis *et al*., 2013). Similarly, other Ts-microRNAs such as miR-145, miR-183, miR-296, miR-96 and miR-126, miR-195, miR-30c, miR-150, and miR-1 were revealed to aberrantly modulate VEGF and N-Ras, villin 2, HMGA1, K-Ras, E2F3 and Cyclin D3, MTA1, Mucin 4, and purine nucleoside phosphorylase (PNP) respectively in many cancer cell lines and *in-vivo* (Grammatikakis *et al*., 2013).

Overall alterations in microRNA regulated oncogenes and tumour suppressor genes result in cancer progression, proliferation and metastasis. While under basal conditions expression of microRNAs acting as oncogenes gets up-regulated whereas those acting as tumour suppressor genes are down–regulated under cancer conditions. The same microRNAs can function either as oncogenes or as tumour suppressor

genes depending on the cellular context. Hence, understanding the molecular mechanisms of microRNAs is of utmost importance in cancer prediction, diagnosis and therapy.

1.13 MicroRNAs in breast cancer

In BC, several microRNAs have been identified as critical regulators of tumour initiation, metastasis and chemo resistance (Table 1.1).

Table 1.1: microRNAs in breast cancer (Adapted from Takahashi *et al.*, 2015).

Phenotype		miRNA	Target Genes
Tumor initiation	Inhibition of self-renewal activity and de-differentiation	let-7	*RAS, HMGA2*
		miR-200c	*BMI-1*
	EMT	miR-200 family & miR-205	*ZEB1*
		miR-103/107	*DICER*
		miR-22	TET family (*TET1-3*)
	Drug resistance	miR-451	*ABCB1*
		miR-326	*ABCC1*
		miR-487a	*ABCG2*
		miR-221/222	$p27^{kip1}$
		miR-30c	*TWF1* and *IL-11*
		miR-31	*PKCepsilon*
	Invasion and metastasis	miR-10b	*HOXD1*
		miR-335	*SOX4, TNC*
		miR-31	*RhoA & ITGA5*
		miR-34	*Snail*
		miR-29b	*VEGFA, ANGPTL4, LOX*
		miR-708	*NNAT*

Abbreviations: ANGPTL4, angiopoietin-like 4; BMI-1, B cell-specific Moloney murine leukemiavirus integration site 1; EMT, epithelial to mesenchymal transition; HOXD1, Homeobox D1; IL-11, interleukin-11; ITGA5, integrin 5; LOX, lysyl oxidase; PKCepsilon, protein kinase C epsilon; TNC, tenascin C; TWF1, twinfilin 1; VEGFA, vascular endothelial growth factor A.

Breast cancer initiation has been attributed in part to the cancer initiating cells or cancer stem cells (CSC) regulated by several microRNAs. It has reported that CSCs can undergo asymmetric cell division with synthesis and release of small compounds similar to somatic stem cells. Furthermore,various studies reported that these CSCs significantly down-regulate let-7 expression which inhibits the self-renewal and de-differentiation capability of BC cells through RAS and HMGA2 respectively (Yu *et al*., 2007). These cell lineages also display the features of progressive tumourigenicity and epithelial to mesenchymal transition (EMT) (Mani *et al*., 2008). EMT was reported to be correlated with several microRNA family members like miR-205 and miR-200 family (miR-200a, miR-200b, miR-200c, miR-141 and miR-429), in BC cells EMT phenotype is induced by miR-103/107, which down regulates the overall microRNAs synthesis through attenuation of gene encoding Dicer (Martello *et al*., 2010). Moreover, miR-200 family members were found to be regulated via miR-22-mediated suppression of Ten Eleven Translocation (TET) family members (TET1-3). TET family members stimulate miR-200 promoter demethylation resulting in development of CSC properties, such as EMT and metastasis (Song *et al*., 2013).

Several microRNAs have been attributed to drug resistance in BC. Multiple studies have revealed that BC CSCs overexpress ABCC1, also known as MRP1 and ABCG2, the ATP-binding cassette (ABC) transporters responsible for efflux of anti-cancer agents (Sullivan *et al*., 2000; Kim *et al.*, 2001; Scheffer *et al*., 2002; Ross, 2000). These ABC transporters are regulated by several microRNAs at post transcriptional level. In BC, miR-451, miR-326, and miR-487a enhance the chemosensitivity of cells

by regulating ABC transporters activity (Kovalchuk *et al*., 2008; Liang *et al*., 2010; Ma *et al*., 2013). Recent studies also demonstrated the role of another microRNA family (miR-221/222 and miR-30c) in BC drug resistance via modulation of nucleolin and twinfilin 1 respectively (Pichiorri *et al*., 2013; Bockhorn *et al*., 2013).

Apart from drug resistance, microRNAs are also depicted in metastatic processes in BC. MiR-10 has been implicated as a key regulator in BC metastasis (Ma *et al*., 2007). MicroRNAs critically regulate the initial step of metastasis as several microRNAs (miR-126, miR-206, miR-335, miR-31, miR-34a, miR-22, miR-200) family,(miR-29b, and miR-708) inhibited BC metastasis (Tavazoie *et al*., 2008; Chou *et al*., 2013; Ryu *et al*., 2013). In-addition, exosomal microRNAs have also been implicated in BC invasion and metastasis (Zhou *et al*., 2014). Exosomes are membrane vesicles of a 50–100 nm size derived from various cell types and detected in biologic fluids in both physiological and pathological conditions (Taylor *et al*., 2013).

Studie revealed that bone marrow stroma derived exosomal microRNAs (miR-127, miR-197, miR-222, and miR-223) promotes long term survival of cancer cells by inhibiting proliferation and switching cells to a dormant stage which recurred years after disease diagnosis and treatment (Lim *et al*., 2011). Similarly, other exosomal microRNAs like miR-105 are also reported to be involved in high metastatic activity (Zhou *et al*., 2014).

Although these biomarkers for BC have been identified, there is a need to look for novel microRNAs that may be involved in the early diagnosis of BC.

1.14 Methods for biomarkers discovery

1.14.1 Proteomics

Proteomics can be defined as the study of an entire proteome and protein expression profile of a specific cell type or tissue in a given setting using high throughput technologies (Baak *et al*., 2005). The objective is to identify sets of protein which are differentially expressed between a normal and pathological / disease state (Baak *et al*., 2005) thus potentially creating a unique profile or fingerprint for a particular disease (e.g. BC). In BC, useful biomarkers for BC diagnosis have been identified during such studies (Liu *et al*., 2013). Example identified by this technique are Serum biomarkers to detect BC (BC1, BC2, BC3, and CA 15.3) (Khalilpour *et al*., 2017).

1.14.2 Transcriptomics

Genetic profiling has been used in medical research for the identification of genetic markers required for disease screening including cancer (Davis *et al*., 2006; Tainsky, 2009). DNA microarray is one of the most common high throughput techniques used to study genome wide gene expression profiles in cells or tissues. It has enabled researchers to simultaneously resolve the mRNA expression levels of thousands of genes in an organism (Karakach*et al*., 2010; Pulverer *et al*., 2012). DNA microarrays are collections of microscopic spots created by robotic machines and arranged in a

grid-like format on a solid support such as a glass slide. Each microscopic spot represents cDNA derived from mRNA of known genes. The cDNA of several thousand genes can be spotted in a single process by performing DNA microarray analysis. DNA microarray analysis involves a numerous steps starting with (a) the design of the experiment, (b) extraction of nucleic acids (usually mRNA) from the control (healthy cells or tissue) and experimental (BC cells or tissue) samples, (c) transcription of extracted mRNA into cDNA molecules that are differently labelled (the controlled sample is labelled with cy3) with fluorescent labels, (d) hybridization of labelled cDNA molecules with cDNA immobilized on glass slides, (e) scanning of microarray, (f) image processing, (g) normalized ratio calculation, (h) statistical analysis, and (i) concluding with information extraction and generation of knowledge from results (Karakach *et al*., 2010).

1.15 Bioinformatics as a tool for the detection of novel biomarkers

Bioinformatics is the application of computer technology in the management of biological information. In 1979, Paulien Hogeweg coined the term bioinformatics for the study of informatics processes in biotic systems (Raza *et al*., 2012). It is concerned with storage, extracting, organizing, interpreting and utilizing information from biological sequences and molecules (Baxevanis *et al*., 2004). The main aim of bioinformatics is to find key biological information hidden amongst a mass of raw data to identify important trends and patterns which would eventually lead to novel biomarker discovery for both diagnostic and therapeutic purposes (Raza, 2012).

However, to achieve this aim three crucial steps are required. The first is accessibility of data, i.e. data needs to be ordered in such a way that users are able to access existing data as well as submit new findings. Secondly to develop tools to facilitate data analysis and lastly to use these tools to analyse and interpret results in such a way that they become biologically significant (Luscombe *et al*., 2001; Wu *et al*., 2012). With advancements in computational technology, bioinformatics continued to progress towards the production of specialized automated systems, algorithms, databases and software to manage the sheer volume of sequence data generated from growing collection of human genome sequences and germ line polymorphisms (Raza *et al*., 2012). Furthermore, bioinformatics has presented ways in which data mining approaches can be used to filter valuable targets such as microRNAs, genes, or proteins for the discovery of possible novel biomarkers for diseases (Akhtar *et al*., 2016).

In diseases like cancer, where the affected cells are arranged in complex or even unpredictable ways, one of the ways to apply bioinformatics methods with regard to signalling, proliferation, communication and specificity of disease metabolism is through cancer bioinformatics (Wu *et al*., 2012). Using cancer bioinformatics, biomarkers for specific clinical phenotypes with respect to early diagnosis, disease prediction and treatment response can be identified and validated to improve patient's quality of life (Wu *et al*., 2012). Furthermore with the advent of cancer bioinformatics it has now become possible to study dynamic network of biomarkers evolved through integration of gene-gene, gene-protein or protein-protein interactions

at different disease stages and time points. Another developing science is clinical bioinformatics, which merges medical informatics, clinical informatics, bioinformatics, mathematics, omics science and information technology together. Clinical bioinformatics is considered as one of the key factors for addressing important clinical challenges in early diagnosis, predictive prognosis and effective therapies in cancer patients (Wu *et al*., 2012). It is urgently required to correlate the outcomes of cancer bioinformatics with clinical bioinformatics, where gene/protein interactions can be studied with respect to patient's complaints, history, therapies, clinical symptoms and signs, physician's examinations, biochemical analyses, imaging profiles, pathologies and other measurements (Wang, 2011). However, this association seems to have a number of challenges like validation of accuracy and sensitivity of integrated systems developed to translate the clinical information to clinical informatics, bioinformatics analysis with regard to disease severity, duration, location, drug treatment, or computational integration of all elements for a precision conclusion.

1.16 Aims of thesis

The overall aim of this thesis was to identity novel microRNAs as biomarkers for early stage detection of BC.

The specific objective were:

i). To identify microRNA and their gene targets for BC diagnosis using several bioinformatics tools.

ii). To identify the prognostic and predictive values of the novel microRNAs and their target genes using a bioinformatics approach.

iii). To determine the expression levels of the identified microRNA and their gene targets in BC tissues as well as different cancer tissues using several bioinformatics tools.

1.17 References

Abdelmohsen, K. and Gorospe, M., 2010. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdisciplinary Reviews: RNA*, 1(2), pp.214-229.

Abdelmohsen, K., Kim, M.M., Srikantan, S., Mercken, E.M., Brennan, S.E., Wilson, G.M., de Cabo, R. and Gorospe, M., 2010. miR-519 suppresses tumor growth by reducing HuR levels. *Cell cycle*, 9(7), pp.1354-1359.

Abdelmohsen, K., Srikantan, S., Tominaga, K., Kang, M.J., Yaniv, Y., Martindale, J.L., Yang, X., Park, S.S., Becker, K.G., Subramanian, M. and Maudsley, S., 2012. Growth inhibition by mir-519 via multiple p21-inducing pathways. *Molecular and Cellular Bology*, 32(13), pp.2530-2548.

Acunzo, M., Romano, G., Wernicke, D. and Croce, C.M., 2015. MicroRNA and cancer–a brief overview. *Advances in biological regulation*, 57, pp.1-9.

Akhtar, M.M., Micolucci, L., Islam, M.S., Olivieri, F. and Procopio, A.D., 2016. Bioinformatic tools for microRNA dissection. *Nucleic acids research*, 44(1), pp.24-44.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J. and Clarke, M.F., 2003. Prospective identification of tumorigenic breast cancer cells. Proceedings of the *National Academy of Sciences*, 100(7), pp.3983-3988.

Alobaidi, A.H., Jalaly, A. and Alsamarai, A.M., 2015. Biomarkers in Women with Breast Cancer: I: CEA, CA 15.3, CA 27-29, BRCA1, and BRCA2 Predictive Value. *IJSRSET*, 1, pp.342-349.

Alvarez-Garcia, I. and Miska, E.A., 2005. MicroRNA functions in animal development and human disease. *Developmen*t, 132(21), pp.4653-4662.

American Society 2014 Cancer facts & Figure 2014. Accessed at http://www.cancer.org/acs/groups/cid/documents/webcontent/002550- pdfon 13/4/2016.

American Society 2015 Cancer facts & Figure 2015. Accessed at. http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdfon 5/4/2016.

American Society, 2016 Cancer facts & Figure 2016. Accessed at http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf on 10/4/2016.

Arpino, G., Bardou, V.J., Clark, G.M. and Elledge, R.M., 2004. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast cancer research*, 6(3), p.R149.

Arps, D.P., Healy, P., Zhao, L., Kleer, C.G. and Pang, J.C., 2013. Invasive ductal carcinoma with lobular features: a comparison study to invasive ductal and invasive

lobular carcinomas of the breast. *Breast cancer research and treatment*, 138(3), pp.719-726.

Atchley, D.P., Albarracin, C.T., Lopez, A., Valero, V., Amos, C.I., Gonzalez-Angulo, A.M., Hortobagyi, G.N. and Arun, B.K., 2008. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *Journal of Clinical Oncology*, 26(26), pp.4282-4288.

Atun, R., Jaffray, D.A., Barton, M.B., Bray, F., Baumann, M., Vikram, B., Hanna, T.P., Knaul, F.M., Lievens, Y., Lui, T.Y. and Milosevic, M., 2015. Expanding global access to radiotherapy. *The Lancet Oncology*, 16(10), pp.1153-1186.

Baak, J.P., Janssen, E.A., Soreide, K. and Heikkilae, R., 2005. Genomics and proteomics–the way forward. *Ann Oncol*, 16(S2), pp.30-44.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. *cell*, 136(2), pp.215-233.

Baxevanis, A.D. and Ouellette, B.F., 2004. Bioinformatics: a practical guide to the analysis of genes and proteins (Vol. 43). John Wiley & Sons.

Berg, W.A., 2009. Tailored supplemental screening for breast cancer: what now and what next? *American Journal of Roentgenology*, 192(2), pp.390-399.

Bernstein, L., 2002. Epidemiology of endocrine-related risk factors for breast cancer. *Journal of mammary gland biology and neoplasia*, 7(1), pp.3-15.

Bhatt, A.N., Mathur, R., Farooque, A., Verma, A. and Dwarakanath, B.S., 2010. Cancer biomarkers-*Current perspectives*, 132(2), pp 129-149.

Bhattacharya, A., Schmitz, U., Wolkenhauer, O., Schönherr, M., Raatz, Y. and Kunz, M., 2013. Regulation of cell cycle checkpoint kinase WEE1 by miR-195 in malignant melanoma. *Oncogene*, 32(26), pp.3175-3183.

Board, P.C.G.E., 2017. Genetics of Breast and Gynecologic Cancers (PDQ®). Bethesda (MD): National Cancer Institute (US).

Bockhorn, J., Dalton, R., Nwachukwu, C., Huang, S., Prat, A., Yee, K., Chang, Y.F., Huo, D., Wen, Y., Swanson, K.E. and Qiu, T., 2013. MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11. *Nature communications,(* 4), p.1393.

Braithwaite, D., Boffetta, P., Rebbeck, T.R. and Meyskens, F., 2012. Cancer prevention for global health: a report from the ASPO International Cancer Prevention Interest Group. *Cancer Epidemiology and Prevention Biomarkers*, 21(9), pp.1606-1610.

Bravaccini, S., Granato, A.M., Medri, L., Foca, F., Falcini, F., Zoli, W., Ricci, M., Lanzanova, G., Masalu, N., Serra, L. and Buggi, F., 2013. Biofunctional characteristics of in situ and invasive breast carcinoma. *Cellular Oncology*, 36(4), pp.303-310.

Breast Cancer Now 2017. Accessed at http://breastcancernow.org/about-breast-cancer/what-can-cause-breast-cancer/family-history-of-breast-cancer-guide 13/2/2017

Brodersen, J., Jorgensen, K.J. and Gotzsche, P.C., 2010. The benefits and harms of screening for cancer with a focus on breast screening. *Polskie Archiwum Medycyny Wewnętrznej*, 120(3), pp.89-94.

Brünner, N., 2009. What is the difference between "predictive and prognostic biomarkers"? Can you give some examples. *Connection*, 13, p.18.

Buchholz, T.A., 2009. Radiation therapy for early-stage breast cancer after breast-conserving surgery. *New England Journal of Medicine*, 360(1), pp.63-70.

Butz, H., Likó, I., Czirjak, S., Igaz, P., Khan, M.M., Zivkovic, V., Bálint, K., Korbonits, M., Rácz, K. and Patócs, A., 2010. Down-regulation of Wee1 kinase by a specific subset of microRNA in human sporadic pituitary adenomas. *The Journal of Clinical Endocrinology & Metabolism*, 95(10), pp.E181-E191.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. and Croce, C.M., 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National academy of Sciences of the United States of America*, 101(9), pp.2999-3004.

Cancer Association of South Africa (2015) Fact Sheet Breast Cancer in Women "Accessed at. http://www.cansa.org.za/files/2016/09/Fact-Sheet-Breast-Cancer-Women-NCR-2011-web-Sept-2016.pdf on 15/4/2016.

Cancer Association of South Africa. (2015). Fact sheet the top ten cancers per population group. Accessed at http://www.cansa.org.za/files/2015/05/Fact-Sheet-Top-Ten-Cancers-per-Population-Group-May-2015.pdf on 12/8/2016

Cao, A.Y., Huang, L., Wu, J., Lu, J.S., Liu, G.Y., Shen, Z.Z., Shao, Z.M. and Di, G.H., 2012. Tumor characteristics and the clinical outcome of invasive lobular carcinoma compared to infiltrating ductal carcinoma in a Chinese population. *World journal of surgical oncology*, 10(1), p.152.

Castro, E. and Eeles, R., 2012. The role of BRCA1 and BRCA2 in prostate cancer. *Asian journal of andrology*, 14(3), p.409.

Cavalli, F., 2006. Cancer in the developing world: can we avoid the disaster? Nature clinical practice *Oncology*, 3(11), pp.582-583.

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

Chen, D., Farwell, M.A. and Zhang, B., 2010. MicroRNA as a new player in the cell cycle. *Journal of cellular physiology*, 225(2), pp.296-301.

Chen, F. and Hu, S.J., 2012. Effect of microRNA-34a in cell cycle, differentiation, and apoptosis: A review. *Journal of biochemical and molecular toxicology*, 26(2), pp.79-86.

Chou, J., Lin, J.H., Brenot, A., Kim, J.W., Provot, S. and Werb, Z., 2013. GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nature cell biology*, 15(2), pp.201-213.

Chung, L., Moore, K., Phillips, L., Boyle, F.M., Marsh, D.J. and Baxter, R.C., 2014. Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. *Breast cancer research*, 16(3), p.R63.

Cole, M.D. and McMahon, S.B., 1999. The Myc oncoprotein: a critical evaluation of transactivation and target gene regulation. *Oncogene*, 18(19).

Collaborative Group on Hormonal Factors in Breast Cancer, 1997. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *The Lancet*, 350(9084), pp.1047-1059.

Collins, K.K., Liu, Y., Schootman, M., Aft, R., Yan, Y., Dean, G., Eilers, M. and Jeffe, D.B., 2011. Effects of breast cancer surgery and surgical side effects on body image over time. *Breast cancer research and treatment*, 126(1), pp.167-176.

Connor, C. and Attai, D., 2013. Adjuvant endocrine therapy for the surgeon: options, side effects, and their management. *Annals of surgical oncology*, 20(10), pp.3188-3193.

Cook, A.C., Tuck, A.B., McCarthy, S., Turner, J.G., Irby, R.B., Bloom, G.C., Yeatman, T.J. and Chambers, A.F., 2005. Osteopontin induces multiple changes in gene expression that reflect the six "hallmarks of cancer" in a model of breast cancer progression. *Molecular carcinogenesis*, 43(4), pp.225-236.

Coopey, S.B., Tang, R., Lei, L., Freer, P.E., Kansal, K., Colwell, A.S., Gadd, M.A., Specht, M.C., Austen, W.G. and Smith, B.L., 2013. Increasing eligibility for nipple-sparing mastectomy. *Annals of surgical oncology*, 20(10), pp.3218-3222.

Coussens, L., 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science*, 23, pp.1132-1140.

Darby, S.C., Ewertz, M., McGale, P., Bennet, A.M., Blom-Goldman, U., Brønnum, D., Correa, C., Cutter, D., Gagliardi, G., Gigante, B. and Jensen, M.B., 2013. Risk of ischemic heart disease in women after radiotherapy for breast cancer. *New England Journal of Medicine*, 368(11), pp.987-998.

Davis, M.A. and Hanash, S., 2006. High-throughput genomic technology in research and clinical management of breast cancer. Plasma-based proteomics in early detection and therapy. *Breast Cancer Research*, 8(6), p.217.

Dellapasqua, S., Colleoni, M., Gelber, R.D. and Goldhirsch, A., 2005. Adjuvant endocrine therapy for premenopausal women with early breast cancer. *Journal of clinical oncology*, 23(8), pp.1736-1750.

Derveaux, S., Vandesompele, J. and Hellemans, J., 2010. How to do successful gene expression analysis using real-time PCR. *Methods*, 50(4), pp.227-230.

Di Leva, G., Garofalo, M. and Croce, C.M., 2014. MicroRNAs in cancer. Annual Review of Pathology*: Mechanisms of Disease*, 9, pp.287-314.

Di Leva, G., Garofalo, M. and Croce, C.M., 2014. MicroRNAs in cancer. Annual Review of Pathology: *Mechanisms of Disease*, 9, pp.287-314.

Dixon, J.M., Renshaw, L., Bellamy, C., Stuart, M., Hoctin-Boes, G. and Miller, W.R., 2000. The effects of neoadjuvant anastrozole (Arimidex) on tumor volume in postmenopausal women with breast cancer: a randomized, double-blind, single-center study. *Clinical Cancer Research*, 6(6), pp.2229-2235.

Donepudi, M.S., Kondapalli, K., Amos, S.J. and Venkanteshan, P., 2014. Breast cancer statistics and markers. *Journal of cancer research and therapeutics*, 10(3), p.506.

Donzelli, S., Fontemaggi, G., Fazi, F., Di Agostino, S., Padula, F., Biagioni, F., Muti, P., Strano, S. and Blandino, G., 2012. MicroRNA-128-2 targets the transcriptional repressor E2F5 enhancing mutant p53 gain of function. *Cell Death & Differentiation*, 19(6), pp.1038-1048.

Duffy, M.J., 2006. Serum tumor markers in breast cancer: are they of clinical value?. *Clinical chemistry*, 52(3), pp.345-351.

Elmore, S., 2007. Apoptosis: a review of programmed cell death. *Toxicologic pathology*, 35(4), pp.495-516.

Espina, V. and Liotta, L.A., 2011. What is the malignant nature of human ductal carcinoma in situ?. *Nature Reviews Cancer*, 11(1), pp.68-75.

Evans, W., 2012. Breast cancer screening: successes and challenges. CA: a cancer *journal for clinicians*, 62(1), pp.5-9.

Ezzati, M., Lopez, A.D., Rodgers, A., Vander Hoorn, S., Murray, C.J. and Comparative Risk Assessment Collaborating Group, 2002. Selected major risk factors and global and regional burden of disease. *The Lancet*, 360(9343), pp.1347-1360.

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

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*, 136(5), pp.E359-E386.

Fernandez, S.V., Robertson, F.M., Pei, J., Aburto-Chumpitaz, L., Mu, Z., Chu, K., Alpaugh, R.K., Huang, Y., Cao, Y., Ye, Z. and Cai, K.Q., 2013. Inflammatory breast cancer (IBC): clues for targeted therapies. *Breast cancer research and treatment*, 140(1), pp.23-33.

Floyd, D.H., Zhang, Y., Dey, B.K., Kefas, B., Breit, H., Marks, K., Dutta, A., Herold-Mende, C., Synowitz, M., Glass, R. and Abounader, R., 2014. Novel anti-apoptotic microRNAs 582-5p and 363 promote human glioblastoma stem cell survival via direct inhibition of caspase 3, caspase 9, and Bim. *PLoS One*, 9(5), p.e96239..

Fong, K.M. and Minna, J.D., 2002. Molecular biology of lung cancer: clinical implications. *Clinics in chest medicine*, 23(1), pp.83-101.

Frezzetti, D., De Menna, M., Zoppoli, P., Guerra, C., Ferraro, A., Bello, A.M., De Luca, P., Calabrese, C., Fusco, A., Ceccarelli, M. and Zollo, M., 2011. Upregulation of miR-21 by Ras in vivo and its role in tumor growth. *Oncogene*, 30(3), pp.275-286.

Friebel, T.M., Domchek, S.M. and Rebbeck, T.R., 2014. Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: A systematic review and meta-analysis. *Journal of the National Cancer Institute*, 106(6), p.dju091.

Frixa, T., Donzelli, S. and Blandino, G., 2015. Oncogenic microRNAs: Key players in malignant transformation. *Cancers*, 7(4), pp.2466-2485.

Fulda, S., 2009. Tumor resistance to apoptosis. *International Journal of Cancer*, 124(3), pp.511-515.

Ganci, F., Sacconi, A., Ben-Moshe, N.B., Manciocco, V., Sperduti, I., Strigari, L., Covello, R., Benevolo, M., Pescarmona, E., Domany, E. and Muti, P., 2013. Expression of TP53 mutation-associated microRNAs predicts clinical outcome in head and neck squamous cell carcinoma patients. *Annals of oncology*, 24(12), pp.3082-3088.

Garraway, L.A. and Lander, E.S., 2013. Lessons from the cancer genome. *Cell*, 153(1), pp.17-37.

Garrison, L.P., Lalla, D., Brammer, M.G., Wang, B., Babigumira, J. and Perez, E.A., 2011. Assessing the potential cost-effectiveness of retesting IHC0, IHC1-positive, or FISH-negative early-stage breast cancer patients for HER2 status. *Journal of Clinical Oncology*, 29(15_suppl), pp.6133-6133.

Gartel, A.L. and Radhakrishnan, S.K., 2005. Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer research*, 65(10), pp.3980-3985.

Gaudet, M.M., Gapstur, S.M., Sun, J., Diver, W.R., Hannan, L.M. and Thun, M.J., 2013. Active smoking and breast cancer risk: original cohort data and meta-analysis. *Journal of the National Cancer Institute*, p.djt023.

Gerhardus, A., Schleberger, H., Schlegelberger, B. and Gadzicki, D., 2007. Diagnostic accuracy of methods for the detection of BRCA1 and BRCA2 mutations: a systematic review. *European Journal of Human Genetics*, 15(6), pp.619-627.

Gonzaga, M.A., 2010. How accurate is ultrasound in evaluating palpable breast masses*?*. *Pan African Medical Journal*, 7(1).

Grammatikakis, I., Gorospe, M. and Abdelmohsen, K., 2013. Modulation of cancer traits by tumor suppressor microRNAs. *International journal of molecular sciences*, 14(1), pp.1822-1842.

Guessous, F., Zhang, Y., Kofman, A., Catania, A., Li, Y., Schiff, D., Purow, B. and Abounader, R., 2010. microRNA-34a is tumor suppressive in brain tumors and glioma stem cells. *Cell cycle*, 9(6), pp.1031-1036.

Gutschner, T. and Diederichs, S., 2012. The hallmarks of cancer: a long non-coding RNA point of view. *RNA biology*, 9(6), pp.703-719.

Hanahan, D. and Folkman, J., 1996. Patterns and emerging mechanisms of the angiogenic switch during genesis. cell, 86(3), pp.353-364.

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

Harris, S.L. and Levine, A.J., 2005. The p53 pathway: positive and negative feedback loops. *Oncogene*, 24(17), pp.2899-2908.

Hashim, D., Boffetta, P., La Vecchia, C., Rota, M., Bertuccio, P., Malvezzi, M. and Negri, E., 2016. The global decrease in cancer mortality: trends and disparities. *Annals of Oncology*, 27(5), pp.926-933.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. and Hammond, S.M., 2005. A microRNA polycistron as a potential human oncogene. *Nature*, 435(7043), pp.828-833.

Hickey, M., Peate, M., Saunders, C.M. and Friedlander, M., 2009. Breast cancer in young women and its impact on reproductive function. *Human reproduction update*, 15(3), pp.323-339.

Hu, W., Chan, C.S., Wu, R., Zhang, C., Sun, Y., Song, J.S., Tang, L.H., Levine, A.J. and Feng, Z., 2010. Negative regulation of tumor suppressor p53 by microRNA miR-504. *Molecular cell*, 38(5), pp.689-699.

Hudson, R.S., Yi, M., Esposito, D., Glynn, S.A., Starks, A.M., Yang, Y., Schetter, A.J., Watkins, S.K., Hurwitz, A.A., Dorsey, T.H. and Stephens, R.M., 2013. MicroRNA-106b-25 cluster expression is associated with early disease recurrence and targets caspase-7 and focal adhesion in human prostate cancer. *Oncogene*, 32(35), pp.4139-4147.

Hulka, B.S. and Moorman, P.G., 2001. Breast cancer: hormones and other risk factors. *Maturitas*, 38(1), pp.103-113.

Hurst, D.R., Edmonds, M.D. and Welch, D.R., 2009. Metastamir: the field of metastasis-regulatory microRNA is spreading. *Cancer research*, 69(19), pp.7495-7498.

Ivanovska, I., Ball, A.S., Diaz, R.L., Magnus, J.F., Kibukawa, M., Schelter, J.M., Kobayashi, S.V., Lim, L., Burchard, J., Jackson, A.L. and Linsley, P.S., 2008. MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Molecular and cellular biology*, 28(7), pp.2167-2174.

Jafri, M.A., Ansari, S.A., Alqahtani, M.H. and Shay, J.W., 2016. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome medicine*, *8*(1), p.69.

Jemal, A., Bray, F., Forman, D., O'brien, M., Ferlay, J., Center, M. and Parkin, D.M., 2012. Cancer burden in Africa and opportunities for prevention. *Cancer*, 118(18), pp.4372-4384.

Jemal, A., Center, M.M., DeSantis, C. and Ward, E.M., 2010. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiology and Prevention *Biomarkers*, 19(8), pp.1893-1907.

Johnson, C.E., Gorringe, K.L., Thompson, E.R., Opeskin, K., Boyle, S.E., Wang, Y., Hill, P., Mann, G.B. and Campbell, I.G., 2012. Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. *Breast cancer research and treatment*, 133(3), pp.889-898.

Karakach, T.K., Flight, R.M., Douglas, S.E. and Wentzell, P.D., 2010. An introduction to DNA microarrays for gene expression analysis. Chemometrics and Intelligent *Laboratory Systems*, 104(1), pp.28-52.

Kelly, K.M., Dean, J., Comulada, W.S. and Lee, S.J., 2010. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology*, 20(3), pp.734-742.

Khalilpour, A., Kilic, T., Khalilpour, S., Álvarez, M.M. and Yazdi, I.K., 2017. Proteomic-based biomarker discovery for development of next generation diagnostics. *Applied Microbiology and Biotechnology*, 101(2), pp.475-491.

Killela, P.J., Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L.A., Friedman, A.H., Friedman, H., Gallia, G.L., Giovanella, B.C. and Grollman, A.P., 2013. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proceedings of the National Academy of Sciences*, *110*(15), pp.6021-6026.

Kim, R., Nishimoto, N., Osaki, A., Kojima, J., Fujii, S., Inai, K. and Toge, T., 2001. An analysis of relapsed breast cancer in patients previously treated with breast conserving surgery. *Breast cancer*, 8(2), p.146.

Kotsopoulos, J., Lubinski, J., Salmena, L., Lynch, H.T., Kim-Sing, C., Foulkes, W.D., Ghadirian, P., Neuhausen, S.L., Demsky, R., Tung, N. and Ainsworth, P., 2012. Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Research*, 14(2), p.R42.

Kovalchuk, O., Filkowski, J., Meservy, J., Ilnytskyy, Y., Tryndyak, V.P., Vasyl'F, C. and Pogribny, I.P., 2008. Involvement of microRNA-451 in resistance of the MCF-7 breast cancer cells to chemotherapeutic drug doxorubicin. *Molecular cancer therapeutics*, 7(7), pp.2152-2159.

Kumar, M., Lu, Z., Takwi, A.A.L., Chen, W., Callander, N.S., Ramos, K.S., Young, K.H. and Li, Y., 2011. Negative regulation of the tumor suppressor p53 gene by microRNAs. *Oncogene*, 30(7), pp.843-853.

Lando, M., 2011. Genetic alterations in cervical cancer: consequences for gene regulation and clinical outcome after chemoradiotherapy

Law, K.S., Azman, N., Omar, E.A., Musa, M.Y., Yusoff, N.M., Sulaiman, S.A. and Hussain, N.H.N., 2014. The effects of virgin coconut oil (VCO) as supplementation on quality of life (QOL) among breast cancer patients. *Lipids in health and disease*, 13(1), p.139.

Li, L., Yuan, L., Luo, J., Gao, J., Guo, J. and Xie, X., 2013. MiR-34a inhibits proliferation and migration of breast cancer through down-regulation of Bcl-2 and SIRT1. *Clinical and experimental medicine*, 13(2), pp.109-117.

Li, X.M., Yan, H., Zhou, K.N., Dang, S.N., Wang, D.L. and Zhang, Y.P., 2011. Effects of music therapy on pain among female breast cancer patients after radical mastectomy: results from a randomized controlled trial. *Breast cancer research and treatment*, 128(2), pp.411-419.

Liang, Z., Wu, H., Xia, J., Li, Y., Zhang, Y., Huang, K., Wagar, N., Yoon, Y., Cho, H.T., Scala, S. and Shim, H., 2010. Involvement of miR-326 in chemotherapy resistance of breast cancer through modulating expression of multidrug resistance-associated protein 1. *Biochemical pharmacology*, 79(6), pp.817-824.

Lim, P.K., Bliss, S.A., Patel, S.A., Taborga, M., Dave, M.A., Gregory, L.A., Greco, S.J., Bryan, M., Patel, P.S. and Rameshwar, P., 2011. Gap junction–mediated import of MicroRNA from bone marrow stromal cells can elicit cell cycle quiescence in breast cancer cells. *Cancer research*, 71(5), pp.1550-1560.

Lima, R.T., Busacca, S., Almeida, G.M., Gaudino, G., Fennell, D.A. and Vasconcelos, M.H., 2011. MicroRNA regulation of core apoptosis pathways in cancer. *European journal of cancer*, 47(2), pp.163-174.

Liston, P., Fong, W.G. and Korneluk, R.G., 2003. The inhibitors of apoptosis: there is more to life than Bcl2. *Oncogene*, 22(53), pp.8568-8580.

Liu, N.Q., Dekker, L.J., Stingl, C., Güzel, C., De Marchi, T., Martens, J.W., Foekens, J.A., Luider, T.M. and Umar, A., 2013. Quantitative proteomic analysis of microdissected breast cancer tissues: comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *Journal of proteome research*, 12(10), pp.4627-4641.

Liu, X.L., Peng, C.W., Chen, C., Yang, X.Q., Hu, M.B., Xia, H.S., Liu, S.P., Pang, D.W. and Li, Y., 2011. Quantum dots-based double-color imaging of HER2 positive

breast cancer invasion. *Biochemical and biophysical research communications*, 409(3), pp.577-582.

Lowe, S.W. and Lin, A.W., 2000. Apoptosis in cancer *Carcinogenesis* 21 (3): 485–495. Find this article online.

Lu, Y., Thomson, J.M., Wong, H.Y.F., Hammond, S.M. and Hogan, B.L., 2007. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Developmental biology*, 310(2), pp.442-453.

Lukashchuk, N. and Vousden, K.H., 2007. Ubiquitination and degradation of mutant p53. *Molecular and cellular biology*, 27(23), pp.8284-8295.

Luparia, A., Mariscotti, G., Durando, M., Ciatto, S., Bosco, D., Campanino, P.P., Castellano, I., Sapino, A. and Gandini, G., 2013. Accuracy of tumour size assessment in the preoperative staging of breast cancer: comparison of digital mammography, tomosynthesis, ultrasound and MRI. *La radiologia medica*, 118(7), pp.1119-1136.

Luscombe, N.M., Greenbaum, D. and Gerstein, M., 2001. What is bioinformatics? An introduction and overview. *Yearbook of Medical Informatics*, 1(83-100), p.2.

Ma, L., Teruya-Feldstein, J. and Weinberg, R.A., 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, 449(7163), pp.682-688.

Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., Teruya-Feldstein, J., Reinhardt, F., Onder, T.T., Valastyan, S. and Westermann, F., 2010. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nature cell biology*, 12(3), pp.247-256.

Ma, M.T., He, M., Wang, Y., Jiao, X.Y., Zhao, L., Bai, X.F., Yu, Z.J., Wu, H.Z., Sun, M.L., Song, Z.G. and Wei, M.J., 2013. MiR-487a resensitizes mitoxantrone (MX)-resistant breast cancer cells (MCF-7/MX) to MX by targeting breast cancer resistance protein (BCRP/ABCG2). *Cancer letters*, 339(1), pp.107-115.

MacFarlane, L.A. and R Murphy, P., 2010. MicroRNA: biogenesis, function and role in cancer. *Current genomics*, 11(7), pp.537-561.

Madjar, H., 2010. Role of breast ultrasound for the detection and differentiation of breast lesions. *Breast Care*, 5(2), pp.109-114.

Madu, C.O. and Lu, Y., 2010. Novel diagnostic biomarkers for prostate cancer. *J Cance*r, 1, pp.150-177.

Mahmoud, S.M., Paish, E.C., Powe, D.G., Macmillan, R.D., Grainge, M.J., Lee, A.H., Ellis, I.O. and Green, A.R., 2011. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *Journal of Clinical Oncology*, 29(15), pp.1949-1955.

Malhotra, G.K., Zhao, X., Band, H. and Band, V., 2010. Histological, molecular and functional subtypes of breast cancers. *Cancer biology & therapy*, 10(10), pp.955-960.

Malumbres, M. and Barbacid, M., 2001. Milestones in cell division: to cycle or not to cycle: a critical decision in cancer. *Nature Reviews Cancer*, 1(3), pp.222-231.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M. and Campbell, L.L., 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4), pp.704-715.

Marasa, B.S., Srikantan, S., Martindale, J.L., Kim, M.M., Lee, E.K., Gorospe, M. and Abdelmohsen, K., 2010. MicroRNA profiling in human diploid fibroblasts uncovers miR-519 role in replicative senescence. *Aging (Albany NY)*, 2(6), pp.333-343.

Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T. and Parenti, A.R., 2010. A MicroRNA targeting dicer for metastasis control. *Cell*, 141(7), pp.1195-1207.

Mattie, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., Scott, G.K., Fedele, V., Ginzinger, D., Getts, R. and Haqq, C., 2006. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Molecular cancer*, 5(1), p.24.

Mayr, C. and Bartel, D.P., 2009. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138(4), pp.673-684.

McAuliffe, P., 2016. Ductal Carcinoma *In Situ. In Breast Disease* (pp. 131-143). Springer International Publishing.

McPherson, K., Steel, C. and Dixon, J.M., 2000. Breast cancer-epidemiology, risk factors, and genetics. *BMJ: British Medical Journal*, 321(7261), p.624.

Mees, C., Nemunaitis, J. and Senzer, N., 2009. Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer gene therapy*, 16(2), pp.103-112.

Mehta, R.J., Jain, R.K., Leung, S., Choo, J., Nielsen, T., Huntsman, D., Nakshatri, H. and Badve, S., 2012. FOXA1 is an independent prognostic marker for ER-positive breast cancer. *Breast cancer research and treatment*, 131(3), pp.881-890.

Mehta, S., Shelling, A., Muthukaruppan, A., Lasham, A., Blenkiron, C., Laking, G. and Print, C., 2010. Predictive and prognostic molecular markers for cancer medicine. *Therapeutic advances in medical oncology*, 2(2), pp.125-148.

Melnik, B.C., 2015. MiR-21: an environmental driver of malignant melanoma?. *Journal of translational medicine*, 13(1), p.202.

Mestdagh, P., Fredlund, E., Pattyn, F., Schulte, J.H., Muth, D., Vermeulen, J., Kumps, C., Schlierf, S., De Preter, K., Van Roy, N. and Noguera, R., 2010. MYCN/c-MYC-induced microRNAs repress coding gene networks associated with poor outcome in MYCN/c-MYC-activated tumors. *Oncogene*, 29(9), pp.1394-1404.

Mirtavoos-Mahyari, H., Khosravi, A. and Esfahani-Monfared, Z., 2014. Human epidermal growth factor receptor 2 and estrogen receptor status in respect to tumor characteristics in non-metastatic breast cancer. *Tanaffos*, 13(1), p.26.

Mittal, S., Kaur, H., Gautam, N. and Mantha, A.K., 2017. Biosensors for breast cancer diagnosis: A review of bioreceptors, biotransducers and signal amplification strategies. *Biosensors and Bioelectronics*, 88, pp.217-231.

Mohamed, A., Krajewski, K., Cakar, B. and Ma, C.X., 2013. Targeted therapy for breast cancer. *The American journal of pathology*, 183(4), pp.1096-1112.

Mohammed, E.A. and Babiker, A.I., 2014. Investigation of Risk Factors Digestive System Cancer. *articles from the*, p.293.

Moodley, J., Cairncross, L., Naiker, T. and Momberg, M., 2016. Understanding pathways to breast cancer diagnosis among women in the Western Cape Province, South Africa: a qualitative study. *BMJ open*, 6(1), p.e009905.

Nahta, R., Yu, D., Hung, M.C., Hortobagyi, G.N. and Esteva, F.J., 2006. Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. Nature clinical practice *Oncology*, 3(5), pp.269-280.

Nan, Y., Han, L., Zhang, A., Wang, G., Jia, Z., Yang, Y., Yue, X., Pu, P., Zhong, Y. and Kang, C., 2010. MiRNA-451 plays a role as tumor suppressor in human glioma cells. *Brain research*, 1359, pp.14-21.

National Cancer Institute 2016 Stages of Breast Cancer. Accessed at https://www.cancer.gov/types/breast/patient/breast-treatment-pdq#section/_148 on 12/07/2016.

National Cancer Institute 2016 Cancer Statistics. Accessed at https://www.cancer.gov/about-cancer/understanding/statistics.on 14/3/2016.

Nelson, H.D., Tyne, K., Naik, A., Bougatsos, C., Chan, B.K. and Humphrey, L., 2009. Screening for breast cancer: an update for the US Preventive Services Task Force. *Annals of internal medicine*, 151(10), pp.727-737.

Nigg, E.A., 1995. Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays*, 17(6), pp.471-480.

Niikura, N., Kimura, M., Iwamoto, T., Hayashi, N., Shintoku, J., Saito, Y., Suzuki, Y. and Tokuda, Y., 2013. Women prefer adjuvant endocrine therapy to chemotherapy for breast cancer treatment. *Breast Cancer*, 20(1), pp.67-74.

Nijhawan, R.I., Votava, H.J. and Mariwalla, K., 2012. Clinical application and limitations of the fluorescence in situ hybridization (FISH) assay in the diagnosis and management of melanocytic lesions: a report of 3 cases. *Cutis*, 90(4), pp.189-195.

O'donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T., 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, 435(7043), pp.839-843.

Osman, F., Saleh, F., Jackson, T.D., Corrigan, M.A. and Cil, T., 2013. Increased postoperative complications in bilateral mastectomy patients compared to unilateral mastectomy: an analysis of the NSQIP database. *Annals of surgical oncology*, 20(10), pp.3212-3217.

Park, D.I., Yun, J.W., Park, J.H., Oh, S.J., Kim, H.J., Cho, Y.K., Sohn, C.I., Jeon, W.K., Kim, B.I., Yoo, C.H. and Son, B.H., 2006. HER-2/neu amplification is an independent prognostic factor in gastric cancer. *Digestive diseases and sciences*, 51(8), pp.1371-1379.

Park, J.K., Henry, J.C., Jiang, J., Esau, C., Gusev, Y., Lerner, M.R., Postier, R.G., Brackett, D.J. and Schmittgen, T.D., 2011. miR-132 and miR-212 are increased in pancreatic cancer and target the retinoblastoma tumor suppressor. *Biochemical and biophysical research communications*, 406(4), pp.518-523.

Petro-Nustas, W., Tsangari, H., Phellas, C. and Constantinou, C., 2013. Health beliefs and practice of breast self-examination among young Cypriot women. *Journal of Transcultural Nursing*, 24(2), pp.180-188.

Petrucelli, N., Daly, M.B. and Feldman, G.L., 2013. BRCA1 and BRCA2 hereditary breast and ovarian cancer.

Pfaffl, M.W., 2012. Quantification strategies in real-time polymerase chain reaction. Martin Filion, Hg., Quantitative real-time PCR in Applied *Microbiology*, pp.53-62.

Pichiorri, F., Palmieri, D., De Luca, L., Consiglio, J., You, J., Rocci, A., Talabere, T., Piovan, C., Lagana, A., Cascione, L. and Guan, J., 2013. In vivo NCL targeting affects breast cancer aggressiveness through miRNA regulation. *Journal of Experimental Medicine*, pp.jem-20120950.

Pietras, Kristian, and Arne Östman. 2010. Hallmarks of cancer: interactions with the tumorstroma. *Experimental cell research*, 316.(8): 1324-1331.

Pisanò, M., Mezzolla, V., Galante, M.M., Alemanno, G., Manca, C., Lorusso, V., Malvasi, A. and Tinelli, A., 2011. A new mutation of BRCA2 gene in an Italian healthy woman with familial breast cancer history. *Familial cancer*, 10(1), pp.65-71.

Pulverer, W., Weinhaeusel, A., Noehammer, C. and Vierlinger, K., 2012. Principles and Application of Microarray Technology in Thyroid *Cancer Research*. INTECH Open Access Publisher.

Pylayeva-Gupta, Y., Grabocka, E. and Bar-Sagi, D., 2011. RAS oncogenes: weaving a tumorigenic web. *Nature Reviews Cancer*, 11(11), pp.761-774.

Qi, J., Yu, J.Y., Shcherbata, H.R., Mathieu, J., Wang, A.J., Seal, S., Zhou, W., Stadler, B.M., Bourgin, D., Wang, L. and Nelson, A., 2009. microRNAs regulate human embryonic stem cell division. *Cell Cycle*, 8(22), pp.3729-3741.

Radiology Info (2016)" Breast Ultrasound 2014“Accessed at http://www.radiologyinfo.org/en/info.cfm?PG=breastus 13/3/2017.

Radojicic, J., Zaravinos, A., Vrekoussis, T., Kafousi, M., Spandidos, D.A. and Stathopoulos, E.N., 2011. MicroRNA expression analysis in triple-negative (ER, PR and Her2/neu) breast cancer. *Cell cycle*, 10(3), pp.507-517.

Raza, K., 2012. Application of data mining in bioinformatics. *arXiv preprint arXiv:1205.1125*.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *nature*, 403(6772), pp.901-906.

Richie, R.C. and Swanson, J.O., 2003. Breast cancer: a review of the literature. *JOURNAL OF INSURANCE MEDICINE-NEW YORK THEN DENVER*--, 35(2), pp.85-101.

Riley, R.D., Sauerbrei, W. and Altman, D.G., 2009. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. *British journal of cancer*, 100(8), pp.1219-1229.

RISAL DJOHAN, M.D., Gage, E. and Bernard, S., 2008. Breast reconstruction options following mastectomy. *Cleveland Clinic journal of medicine*, 75, p.S17.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A., 2004. Identification of mammalian microRNA host genes and transcription units. *Genome research*, 14(10a), pp.1902-1910.

Romond, E.H., Perez, E.A., Bryant, J., Suman, V.J., Geyer Jr, C.E., Davidson, N.E., Tan-Chiu, E., Martino, S., Paik, S., Kaufman, P.A. and Swain, S.M., 2005. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *New England Journal of Medicine*, 353(16), pp.1673-1684.

Roşian, A.N.C.A., Lazăr, E.L.E.N.A. and Dema, A.L.I.S., 2005. C-erbB2 oncoprotein: prognostic marker in breast cancer. Rom *J Morphol Embryol*, 46, pp.99-104.

Ross, D.D., 2000. Novel mechanisms of drug resistance in leukemia. *Leukemia*, 14(3), p.467.

Ross, J.S., Slodkowska, E.A., Symmans, W.F., Pusztai, L., Ravdin, P.M. and Hortobagyi, G.N., 2009. The HER-2 receptor and breast cancer: ten years of targeted anti–HER-2 therapy and personalized medicine. *The oncologist*, 14(4), pp.320-368.

Ryu, S., McDonnell, K., Choi, H., Gao, D., Hahn, M., Joshi, N., Park, S.M., Catena, R., Do, Y., Brazin, J. and Vahdat, L.T., 2013. Suppression of miRNA-708 by polycomb group promotes metastases by calcium-induced cell migration. *Cancer Cell*, 23(1), pp.63-76.

Sánchez-Lara, K., Morales-Graf, L., Green, D., Sosa-Sánchez, R. and Méndez-Sánchez, N., 2009. Cancer and obesity. *Gaceta medica de Mexico*, 146(5), pp.326-331.

Sapino, A., Goia, M., Recupero, D. and Marchiò, C., 2013. Current challenges for HER2 testing in diagnostic pathology: state of the art and controversial issues. *Frontiers in oncology*, 3, p.129.

Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M.O., Lehman, C.D., Morris, E., Pisano, E., Schnall, M., Sener, S. and Smith, R.A., 2007. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: a cancer journal for clinicians*, 57(2), pp.75-89.

Scheffer, G.L., Pijnenborg, A.C.L.M., Smit, E.F., Müller, M., Postma, D.S., Timens, W., Van Der Valk, P., De Vries, E.G.E. and Scheper, R.J., 2002. Multidrug resistance related molecules in human and murine lung. *Journal of clinical pathology*, 55(5), pp.332-339.

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H. and Jasin, M., 2011. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell*, 145(4), pp.529-542.

Schulte, J.H., Horn, S., Otto, T., Samans, B., Heukamp, L.C., Eilers, U.C., Krause, M., Astrahantseff, K., Klein-Hitpass, L., Buettner, R. and Schramm, A., 2008. MYCN regulates oncogenic MicroRNAs in neuroblastoma. *International journal of cancer*, 122(3), pp.699-704.

Shapiro, C.L. and Recht, A., 2001. Side effects of adjuvant treatment of breast cancer. *New England Journal of Medicine*, 344(26), pp.1997-2008.

Sharan, S.K. and Bradley, A., 1997. MurineBrca2: Sequence, Map Position, and Expression Pattern. *Genomics*, 40(2), pp.234-241.

Shimono, Y., Mukohyama, J., Nakamura, S.I. and Minami, H., 2015. MicroRNA regulation of human breast cancer stem cells. *Journal of clinical medicine*, 5(1), p.2.

Sigal, A. and Rotter, V., 2000. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. *Cancer research*, 60(24), pp.6788-6793.

Simon, H.U., Haj-Yehia, A. and Levi-Schaffer, F., 2000. Role of reactive oxygen species (ROS) in apoptosis induction. *Apoptosis*, 5(5), pp.415-418.

Sitas, F., Parkin, D.M., Chirenje, M., Stein, L., Abratt, R. and Wabinga, H., 2008. Part II: Cancer in Indigenous Africans—causes and control. *The lancet oncology*, 9(8), pp.786-795.

Song, S.J., Poliseno, L., Song, M.S., Ala, U., Webster, K., Ng, C., Beringer, G., Brikbak, N.J., Yuan, X., Cantley, L.C. and Richardson, A.L., 2013. MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell*, 154(2), pp.311-324.

Stefansson, O.A., Jonasson, J.G., Johannsson, O.T., Olafsdottir, K., Steinarsdottir, M., Valgeirsdottir, S. and Eyfjord, J.E., 2009. Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Research*, 11(4), p.R47.

Stratton, M.R., Campbell, P.J. and Futreal, P.A., 2009. The cancer genome. *Nature*, 458(7239), pp.719-724.

Strimbu, K. and Tavel, J.A., 2010. What are biomarkers?. *Current Opinion in HIV and AIDS*, 5(6), p.463.

Sue, G.R., Lannin, D.R., Killelea, B. and Chagpar, A.B., 2013. Predictors of microinvasion and its prognostic role in ductal carcinoma in situ. The American *Journal of Surgery*, 206(4), pp.478-481.

Sullivan, G.F., Yang, J.M., Vassil, A., Yang, J., Bash-Babula, J. and Hait, W.N., 2000. Regulation of expression of the multidrug resistance protein MRP1 by p53 in human prostate cancer cells. *The Journal of clinical investigation*, 105(9), pp.1261-1267.

Sun, L., Wu, Z., Shao, Y., Pu, Y., Miu, W., Yao, J., Wu, Y. and Yang, Z., 2012. MicroRNA-34a suppresses cell proliferation and induces apoptosis in U87 glioma stem cells. *Technology in cancer research & treatment*, 11(5), pp.483-490.

Switzer, N., Lim, A., Du, L., Al-Sairafi, R., Tonkin, K. and Schiller, D., 2015. Case series of 21 patients with extrahepatic metastatic lobular breast carcinoma to the gastrointestinal tract. *Cancer Treatment Communications*, 3, pp.37-43.

Tainsky, M.A., 2009. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1796(2), pp.176-193.

Takahashi, R.U., Miyazaki, H. and Ochiya, T., 2014. The role of microRNAs in the regulation of cancer stem cells. *Alterations of Epigenetics and MicroRNAs in Cancer and Cancer Stem Cell*, p.55.

Takahashi, R.U., Miyazaki, H. and Ochiya, T., 2015. The roles of microRNAs in breast cancer. *Cancers*, 7(2), pp.598-616.

Tanaka, T., Decuzzi, P., Cristofanilli, M., Sakamoto, J.H., Tasciotti, E., Robertson, F.M. and Ferrari, M., 2009. Nanotechnology for breast cancer therapy. *Biomedical microdevices*, 11(1), pp.49-63.

Tanke, H.J., Dirks, R.W. and Raap, T., 2005. FISH and immunocytochemistry: towards visualising single target molecules in living cells. *Current opinion in biotechnology*, 16(1), pp.49-54.

Tao, Z., Shi, A., Lu, C., Song, T., Zhang, Z. and Zhao, J., 2015. Breast cancer: epidemiology and etiology. *Cell biochemistry and biophysics*, 72(2), pp.333-338.

Tarver, T., 2012. Cancer Facts & Figures 2012. American Cancer Society (ACS) Atlanta, GA: American Cancer Society, 2012. 66 p., pdf. Available from.

Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L. and Massagué, J., 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *nature*, 451(7175), pp.147-152.

Taylor, D.D. and Gercel-Taylor, C., 2013. The origin, function, and diagnostic potential of RNA within extracellular vesicles present in human biological fluids.

Tazawa, H., Tsuchiya, N., Izumiya, M. and Nakagama, H., 2007. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proceedings of the National Academy of Sciences*, 104(39), pp.15472-15477.

Tharkar, P., Madani, A.U., Lasham, A., Shelling, A.N. and Al-Kassas, R., 2015. Nanoparticulate carriers: an emerging tool for breast cancer therapy. *Journal of drug targeting*, 23(2), pp.97-108.

Tian, S., Huang, S., Wu, S., Guo, W., Li, J. and He, X., 2010. MicroRNA-1285 inhibits the expression of p53 by directly targeting its 3′ untranslated region. Biochemical and biophysical research communications, 396(2), pp.435-439. Tijsen, A.J., Pinto, Y.M. and Creemers, E.E., 2012. Circulating microRNAs as diagnostic biomarkers for cardiovascular diseases. *American Journal of Physiology-Heart and Circulatory Physiology*, 303(9), pp.H1085-H1095.

Tsang, W.P. and Kwok, T.T., 2008. Let-7a microRNA suppresses therapeutics-induced cancer cell death by targeting caspase-3. *Apoptosis*, 13(10), pp.1215-1222.

Tyrer, J., Duffy, S.W. and Cuzick, J., 2004. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in medicine*, 23(7), pp.1111-1130.

Valastyan, S., Reinhardt, F., Benaich, N., Calogrias, D., Szász, A.M., Wang, Z.C., Brock, J.E., Richardson, A.L. and Weinberg, R.A., 2009. RETRACTED: A

Pleiotropically Acting MicroRNA, miR-31, Inhibits Breast Cancer Metastasis. *Cell*, 137(6), pp.1032-1046.

Valentini, A., Lubinski, J., Byrski, T., Ghadirian, P., Moller, P., Lynch, H.T., Ainsworth, P., Neuhausen, S.L., Weitzel, J., Singer, C.F. and Olopade, O.I., 2013. The impact of pregnancy on breast cancer survival in women who carry a BRCA1 or BRCA2 mutation. *Breast cancer research and treatment*, 142(1), pp.177-185.

Vaz, C., Ahmad, H.M., Bharti, R., Pandey, P., Kumar, L., Kulshreshtha, R. and Bhattacharya, A., 2013. Analysis of the microRNA transcriptome and expression of different isomiRs in human peripheral blood mononuclear cells. *BMC research notes*, 6(1), p.390.

Wang, P., Zhu, C.F., Ma, M.Z., Chen, G., Song, M., Zeng, Z.L., Lu, W.H., Yang, J., Wen, S., Chiao, P.J. and Hu, Y., 2015. Micro-RNA-155 is induced by K-Ras oncogenic signal and promotes ROS stress in pancreatic cancer. *Oncotarget*, 6(25), p.21148.

Wang, X., 2011. Role of clinical bioinformatics in the development of network-based Biomarkers. *Journal of clinical bioinformatics*, 1(1), p.28.

Wang, X.F., Shi, Z.M., Wang, X.R., Cao, L., Wang, Y.Y., Zhang, J.X., Yin, Y., Luo, H., Kang, C.S., Liu, N. and Jiang, T., 2012. MiR-181d acts as a tumor suppressor in glioma by targeting K-ras and Bcl-2. *Journal of cancer research and clinical oncology*, 138(4), pp.573-584.

Warner, E., 2011. Breast-cancer screening. *New England Journal of Medicine*, 365(11), pp.1025-1032.

Watanabe, S., Ueda, Y., Akaboshi, S.I., Hino, Y., Sekita, Y. and Nakao, M., 2009. HMGA2 maintains oncogenic RAS-induced epithelial-mesenchymal transition in human pancreatic cancer cells. *The American journal of pathology*, 174(3), pp.854-868.

Whelan, T.J., Pignol, J.P., Levine, M.N., Julian, J.A., MacKenzie, R., Parpia, S., Shelley, W., Grimard, L., Bowen, J., Lukka, H. and Perera, F., 2010. Long-term results of hypofractionated radiation therapy for breast cancer. *New England Journal of Medicine*, 362(6), pp.513-520.

Witsch, E., Sela, M. and Yarden, Y., 2010. Roles for growth factors in cancer progression. *Physiology*, 25(2), pp.85-101.

Wong, R.S., 2011. Apoptosis in cancer: from pathogenesis to treatment. *Journal of Experimental & Clinical Cancer Research*, 30(1), p.87.

Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D. and Fields, P., 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*, 265(5181), pp.2088-2091.

Worringer, K.A., Rand, T.A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D. and Yamanaka, S., 2014. The let-7/LIN-41 pathway

regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. *Cell stem cell*, 14(1), pp.40-52.

Wu, D., Rice, C.M. and Wang, X., 2012. Cancer bioinformatics: A new approach to systems clinical medicine. *BMC bioinformatics*, 13(1), p.71.

Xia, T., Liao, Q., Jiang, X., Shao, Y., Xiao, B., Xi, Y. and Guo, J., 2014. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. *Scientific reports*, 4, p.6088.

Xu, C.X., Xu, M., Tan, L., Yang, H., Permuth-Wey, J., Kruk, P.A., Wenham, R.M., Nicosia, S.V., Lancaster, J.M., Sellers, T.A. and Cheng, J.Q., 2012. MicroRNA miR-214 regulates ovarian cancer cell stemness by targeting p53/Nanog. *Journal of Biological Chemistry*, 287(42), pp.34970-34978.

Xue, H., Lu, B. and Lai, M., 2008. The cancer secretome: a reservoir of biomarkers. *Journal of translational medicine*, 6(1), p.52.

Yang, J., Liu, X., Bhalla, K., Kim, C.N., Ibrado, A.M., Cai, J., Peng, T.I., Jones, D.P. and Wang, X., 1997. Prevention of apoptosis by Bcl-2: release of cytochrome c from mitochondria blocked. *Science*, 275(5303), pp.1129-1132.

Yang, Y. and Wang, H., 2013. Perspectives of nanotechnology in minimally invasive therapy of breast cancer. *Journal of healthcare engineering*, 4(1), pp.67-86.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J. and Song, E., 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell*, 131(6), pp.1109-1123.

Zenz, T., Mohr, J., Eldering, E., Kater, A.P., Bühler, A., Kienle, D., Winkler, D., Dürig, J., van Oers, M.H., Mertens, D. and Döhner, H., 2009. miR-34a as part of the resistance network in chronic *lymphocytic leukemia. Blood*, 113(16), pp.3801-3808.

Zhang, B. and Farwell, M.A., 2008. microRNAs: a new emerging class of players for disease diagnostics and gene therapy. *Journal of cellular and molecular medicine*, 12(1), pp.3-21.

Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A., 2007. microRNAs as oncogenes and tumor suppressors. *Developmental biology*, 302(1), pp.1-12.

Zhang, X., Shi, L., Chen, G. and Yap, Y.L., 2011. Integrative omics technologies in cancer biomarker discovery. *Omics Technol. Cancer Biomark. Discov*, 108(33), pp 13653–13658. doi/10.1073/pnas.1103360108

Zhang, Y., Geng, L., Talmon, G. and Wang, J., 2015. MicroRNA-520g confers drug resistance by regulating p21 expression in colorectal cancer. *Journal of Biological Chemistry*, 290(10), pp.6215-6225.

Zhao, A.H., Zeng, Q., Xie, X.Y., Zhou, J.N., Yue, W., Li, Y.L. (2012).MicroRNA-125b induces cancer cell apoptosis through suppression of Bcl-2 expression. *Journal of. Genet. Genomics*,39: 29–35.

Zhou, W., Fong, M.Y., Min, Y., Somlo, G., Liu, L., Palomares, M.R., Yu, Y., Chow, A., O'Connor, S.T.F., Chin, A.R. and Yen, Y., 2014. Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer cell*, 25(4), pp.501-515.

CHAPTER 2

2. *In-silico* approach for identification of microRNAs as biomarkers for breast cancer detection

2.1 Introduction

MicroRNAs are short length RNAs which do not code for proteins; however they play key roles in overall regulation of messenger RNA. MicroRNA partially binds with target mRNAs on its complementary sequences in the 3'UTR or 5'UTR regions and thereby modulate gene expression through mRNA degradation and inhibition or activation of transcriptional processes. In addition, one mRNA can be targeted by multiple microRNAs, and one microRNA can target many mRNAs (Fendler, 2011). A large number of microRNA have been identified using *in vitro* approaches. However, these methods are time-consuming, laborious and expensive with limited utility to predict genome-wide microRNA targeting (Radfar, 2014). In this regard, bioinformatics tools serve as a potential modality to identify new target biomolecule genes of interest. Bioinformatics tools are fast, inexpensive and specific in providing a priority list of potential biomarkers.

To achieve this goal, various target prediction algorithms were developed to identify potential biomarkers. The availability of large sets of sequence data in combination with advanced computational biology offers an optimistic framework for *in silico*

gene expression analysis. However, the availability of large data sets raises its own problems like the presence of a high level of redundancy that was found in the sequences. Therefore, data mining, also known as Knowledge Discovery in Databases (KDD), which involves mining of specific data of interest from a vast amount of data,which plays a crucial role. In bioinformatics, data mining includes the optimization of disease treatment, the function of proteins as well as the prognostics and the diagnostics of diseases (Raza, 2012). Similar to data mining, text mining is also important in assembling and organizing sequence data for *in silico* gene expression analysis. Text mining, which involves extraction of specific information from research articles typically encompass the use of algorithms and statistical information to acquire specific information (Renganathan, 2017).

Each algorithm has its own criteria for binding predictions of microRNAs which include: the free energy of the microRNA, perfection of complementarity of the seed sequence and proximity of different binding sites of the same microRNA (Lekprasert, 2012; Schmitz *et al*., 2014). Although there are many target prediction tools available, challenges regarding standardized methodologies for microRNA target recognition exist. Therefore, the predicted target genes of specific microRNAs as well as miRNAs should be validated using molecular methods for transcripts (qRT-PCR, Northern blotting) or protein products (Western blot, ELISA) (Ji Diana Lee *et al*., 2015).

2.2 Aim and objectives

The aim of this chapter was to identify novel microRNAs that can be used for the early detection of BC. The objectives are to:

i) collect microRNA by employing an *in silico* methodology using miR2disease, GEO, and mirCancer databases;

ii) identify novel microRNAs that were not previously associated with BC through literature mining with the use of databases OncomiRDB, DIANA-TarBase, Pubmed, EXIQON and Google scholar;

iii) predict microRNA target genes using Targetscan and miRDB;

iv) implicate microRNA target genes in cancer related processes through Functional Annotation using DAVID as well as expression analysis using the Human Protein Atlas;

v) identify co-expressed genes of the microRNA target genes using STRING and Gene Expression Atlas;

vi) relate target genes implicated in cancer processes back to individual microRNAs to create a final miRNA list for validation as BC biomarkers.

2.3 MicroRNA databases for breast cancer biomarker discovery

Several databases containing many microRNAs are used for different cancers including BC.

2.3.1 Databases for data mining

2.3.1.1 miR2Disease

miR2Disease, accessible at http://www.mir2disease.org/, is a free online database with specific information on microRNAs for various diseases. Moreover, the data also contains information about the gene expression related to each microRNA as well as an academic reference (Jiang *et al*., 2009). The database can also be searched using search terms like microRNA ID or microRNA disease.

2.3.1.2 miRCancer

miRCancer (available at http://mircancer.ecu.edu/) is another free online database for the identification of microRNAs for different cancers. mirCancer uses text mining to identify microRNAs. The database can also be searched either by the name of the cancer or using a particular microRNA. Furthermore, the database also provides references to articles as well as expression analysis data on each microRNA (Xie *et al*., 2013).

2.3.1.3 Gene expression omnibus (GEO)

Gene expression omnibus (GEO) is an online database (accessible at http://www.ncbi.nlm.nih.gov/projects/geo/) that contains information about gene expression. The database accepts unprocessed or raw data, but usually processed data are submitted to GEO (Barrett & Edgar, 2006). GEO has about 1 billion gene entries

stored derived from many species and cited by many researchers (Barrett & Edgar, 2006)

2.3.1.4 EXIQON

Exiqon contains various databases and allows the use of miRSearch 3.0, an online search tool as part of the Exiqon tool set. Freely available at https://www.exiqon.com/mirsearch, this tool quickly finds and displays specific information on microRNAs relevant to research on diseases with detailed information about each microRNA in the most recent version of miRBase. With the help of a microRNA name the following variables can be searched: a) miRbase history (names and sequences), b) regulated genes (validated and predicted interactions), c) potentially co-transcribed microRNAs, d) diseases in which the microRNA have been shown to be regulated and e) tissues/samples in which the microRNA has been found. Furthermore, validated targets as well as diseases and tissue/sample information is supported by a references with links to PubMed.

2.3.1.5 OncomiRDB

OncomiRDB is an exclusive resource for annotating experimentally verified cancer-related microRNAs with direct evidence. By comparing microRNAs with numerous existing microRNA target databases, oncomiRDB can provide many new microRNA-target annotations validated by a luciferase assay, although it focuses on collecting

cancer related microRNA targets (Wang *et al*., 2014). OncomiRDB can be freely accessed at http://webcache.googleusercontent.com/search?q=cache:http://bioinfo.au.tsinghua.edu.cn/oncomirdb).

2.3.2 Databases for text mining

2.3.2.1 Google Scholar

Google Scholar is a freely accessible website that allows to search across a broad range of academic literature. It draws on information from university repositories, journal publishers, and other websites that has been identified as scholarly by Google Scholar. google scholar is an easy to use website that is comprised of many resources like articles, books and other documents. In November 2013, a feature was added that allows logged-in users to save investigated results into the "Google Scholar library", a personal collection in which the user can search separately and organize the searches by tags (James, 2013).

2.3.2.2 PubMed

PubMed accessible at https://www.ncbi.nlm.nih.gov/pubmed is a free search engine which primarily accesses the MEDLINE database of references. PubMed, the free, home- and office-based MEDLINE ushered in the era of private searching was first released in January 1996. In June 1997, the PubMed system was offered free to the

public by the USA's Vice President, Al Gore when MEDLINE searches via the Web were established during a ceremony (Lindberg, 2000).

2.3.2.3 DIANA-TarBase v7.0

A DIANA-TarBase online bioinformatics tool accessible at (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index) was initially released in 2006. DIANA-TarBase v7.0 was the first to provide hundreds of thousands of high quality manually curated, experimentally validated microRNA gene interactions, which have been enhanced with detailed meta-data (Vlachos *et al*., 2015). With DIANA-TarBase v7.0 one can easily recognize positive or negative experimental results, the utilized experimental methodology, experimental situation as well as cell/tissue type and treatment. The new interface also provides complex information from binding site locations (as identified *in vitro* as well as *in silico*) to primer sequences used for cloning experiments (Vlachos *et al*., 2015). In short, the database allows for identification of microRNA gene targets, and was explored for its literature mining capacity.

2.3.3 Databases for identifying microRNA gene targets

2.3.3.1 TargetScan

TargetScan is used for the prediction of target genes of microRNAs by matching the seed regions of each microRNA with the 8mer and 7mer sites on the target gene (Agarwal *et al*., 2015, Lewis *et al*., 2005). The identification of microRNA by these

sites are also done by matching the seed regions to 3' UTR segments (Friedman *et al*., 2009). In humans, the target prediction rank is dependent on calculating context+ scores of the sites (Grimson *et al.,* 2007). The predict target are also ranked by their probability of conserved targeting (Friedman *et al*., 2009). Target Scan Human uses the miRNA ID as a query to extract target gene information from all publicly available experiments. Targetscan is available at http://www.targetscan.org/vert_71/ (Agarwal *et al.,* 2015).

2.3.3.2 miRDB

miRDB is an online database (accessible at http://mirdb.org) for microRNA target prediction and functional annotations (Wong & Wang, 2015). miRDB hosts predicted microRNA targets for five species: human, mouse, rat, dog and chicken. All the targets in miRDB were predicted using MirTarget, a bioinformatics tool. This tool was developed by analyzing thousands of microRNA-target connections from high-throughput sequencing experiments. Common features linked with microRNA target binding have been recognized and are used to predict microRNA targets through machine learning methods (Wong & Wang, 2015).

2.3.3.3 DAVID functional characterization of genes

DAVID is a bioinformatics database which provides functional interpretation and analysis of a large list of genes from the human, mouse, rat and fly genomes (Dennis

et al., 2003). DAVID is a freely available database that systematically maps a large number of genes to a list of associated Gene Ontology (GO) terms and then statistically highlights gene that are over enriched for those terms as well as for other specific terms (Ashburner *et al*., 2000). This database has four modules: (1) Annotation tools which provide several annotation options, such as GenBank, Unigene, LocusLink, RefSeq, Gene Symbole, GnebName, OMIMm Affymetrix description, Summary and Gene Ontology; each one of these options can be used for a different function depending on the type of analysis. (2) Kyoto Encyclopaedia of Genes and Genomes (KEGG) has charts, which graphically display the distribution of differentially expressed genes in a KEGG biochemical pathway, with each gene linked to the KEGG pathway map. (3) Domain chart displays the distribution of differentially expressed genes and their PFAM protein domains and (4) GO chart, which graphically displays the distribution of differentially expressed genes to functional categories; which consists of three types of classifications (i) Biological Processes, (ii) Molecular Functions and (iii) Cellular Components (Dennis *et al*., 2003, Huang *et al*., .2007).

2.3.4 Databases for protein expression analysis of microRNA gene targets

2.3.4.1 Databases for expression analysis of microRNA gene targets. The Human Proteins Atlas V 15

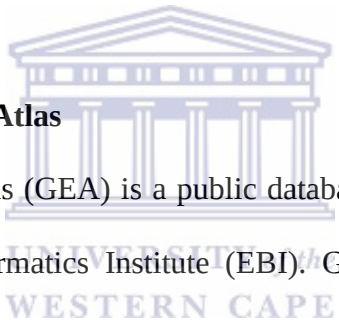
The Human Protein Atlas is a public database that provides the protein expression profiles for a huge amount of normal and cancer tissue proteins from human and is presented as immunohistochemistry images (Persson *et al*., 2006). In 2005, the first

version of the Human Protein Atlas was accessible and it has been followed up by annual releases that have enhanced the coverage of the human proteome. For each new release, the number of antibodies has increased and new structures have provided further functionality (Pontén *et al*., 2011). It is a freely available interactive resource included as part of the Human Protein Atlas portal (www.proteinatlas.org). This site can be helpful for many clinical and biomedical research initiatives including the diagnostic and prognostic value of proteins for many different cancers. The Human Protein Atlas contains numerous data on gene expression and alterations in the expression of these genes and can aid in the discovery of biomarkers for disease by using an *in silico* approach (Uhlen & Fredrik, 2005; Pontèn *et al*., 2008).

2.3.4.2 STRING gene/protein interaction analysis

The STRING version 10 (Search Tool for the Retrieval of Interacting Genes) http://string-db.org/ is an online database that has been designed with the goal of assembling, evaluating and disseminating protein–protein association information in an accessible and complete manner (Szklarczyk *et al*., 2014). STRING was designed on the basis of interactions between proteins as it represents a crucial component of modern biology and hence can be used to detect possible protein interactions between the predicted target genes (Franceschini *et al*., 2013.). The basis of these associations is through data mining of various databases and literature mining throughput, *in vitro* data as well as genomic context analysis prediction (Von Mering *et al*., 2005). STRING integrates and ranks these associations by benchmarking them against a

common reference set and present evidence in a consistent and intuitive web interface. Furthermore, the protein interactions derived from one organism can also be extended to other organisms through automatic transfer of orthologous protein pairs in the test organisms (Von Mering *et al*., 2005). STRING includes a wealth of accessory information, for example protein domains and protein structures, improving its day-to-day value for users. It provides unparalleled comprehensive coverage with 5 million proteins, more than 200 million interactions and more than 1000 organisms (Franceschini *et al*., 2013).



2.3.4.3 Gene Expression Atlas

The Gene Expression Atlas (GEA) is a public database (www.ebi.ac.uk/gxa/) hosted by the European Bioinformatics Institute (EBI). GEA can be used for statistical analysis of data of interest from the Array Express Archive of Functional Genomics Data. GEA allows users to investigate gene expression under several biological and physiological states of different cell types, disease states and developmental stages (Kapushesky *et al*., 2009). Theis database contains information about more than 200,000 genes from nine species and almost 4,500 biological studies using over 30,000 assays. GEA can help investigators determine where a gene of interest can be found in the organism and which genes are differentially expressed in a status or site, for example in a disease or in an organ (Kapushesky *et al*., 2009).

2.4 Methodology

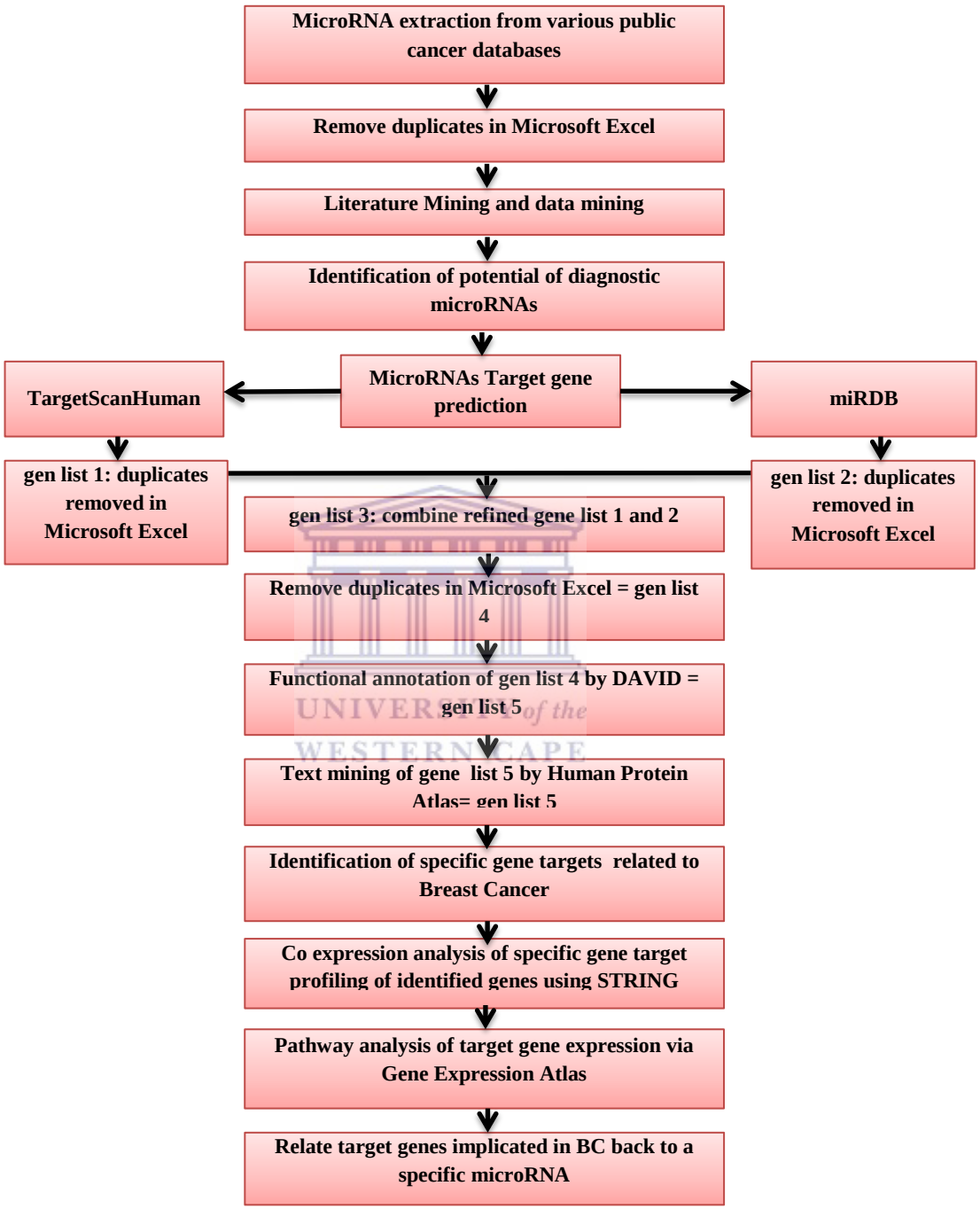


Figure 2.1. Flow diagram showing the *in silico* methodology used for breast cancer microRNA biomarker discovery.

2.5 MicroRNA identification using miR2Disease, GEO and miRCancer databases

2.5.1 miR2Disease

The search for BC microRNAs was initiated using the miR2Disease database see section 2.2.1.1. From the database options, *Homo sapiens* were chosen and a search was performed by using options like; by microRNA, by disease name or by target gene. The resultant list from miR2Disease was saved in a text file.

2.5.2 miRCancer

miRCancer, a microRNA Cancer Association Database was utilized as described in section 2.2.2 to search for by microRNA name or by cancer type. The search was done using the following steps: under *cancer name*, 'breast cancer' was inserted. The search results displayed 333 microRNAs related to BC from which up-regulated microRNAs were chosen and the final results were saved as a text file.

2.5.3 GEO

Next dataset used in the search for microRNAs was GEO, as mentioned in section 2.2.1.3. In the *key words or GEO Accession* field, the key term 'breast cancer microRNAs were inserted and the *search* box was activated. The search results displayed two databases out of which the first database (i.e. GEO Dataset Database) was used. Using GEO Dataset Database, 2799 search results for BC was obtained and

selected. Specific microRNAs for BC was identified by referring to each article available in the database.

2.6 Literature mining for novel microRNA identification

The list of microRNAs, obtained from the initial database screening was further evaluated by literature mining. Literature mining was conducted in order to identify a smaller subset of microRNAs as putative biomarkers for BC. These non-validated microRNA biomarkers for BC were further searched for their novelty by using databases Exiqon, OncomiRDB, Google scholar, Pubmed and Diana-TarBase v7.0 database.

2.6.1 EXIQON

On the homepage of Exiqon, *miRSearch* was selected and 'breast cancer' was searched using the tool '*Find microRNAs relevant for your research'* box. Additionally, '*Human*' was chosen from the '*Organism* ' dropdown menu. The results of this search were saved as a text file.

2.6.2 OncomiRDB

In oncomiRDB database, microRNAs was searched for by *microRNA ID* or *tumor tissue*. 'Breast cancer' was used as a keyword within the *tumor tissue* field, and the results were saved as a text file.

2.6.3 Google Scholar

Google scholar was used to perform a general search on each of the microRNAs (searched for by name) using the combined list generated from the database screening sections 2.5.1 – 2.5.3. In the event that a publication on any given microRNA was discovered implicating it in BC, it would be excluded from the list. The results of this search were saved as a text file.

2.6.4 PubMed

Using PubMed, a general search was performed on each of the microRNAs (searched for by name) using the combined list resulting from the database screening section 2.5.1 – 2.5.3. Already published microRNA data on BC in PubMed was excluded from the analysis. The results of this search were saved as a text file.

2.6.5 Diana database

DIANA is database used as described in section 2.2.2.3, the microRNA ID was used as input to search for the microRNAs already associated with breast cancer within the database which was then excluded from further analysis. The results of this search were saved as a text file.

2.7 Target gene prediction for selected microRNAs

The names of each of the final shortlisted microRNAs from literature mining were used to predict their gene targets using miRDB and TargetScanHuman databases.

2.7.1 miRDB database

Each of the shortlisted microRNAs was initially inserted into the miRDB search box using the '*Human'* option. A list of target genes for each microRNA was obtained from miRDB and each microRNA was given a target prediction score. Only genes with a target prediction score greater than 90 % were included for further analysis. Similar target prediction analysis was also performed using TargetScanHuman.

2.7.2 TargetScanHuman databases

Each of the final listed microRNAs was individually entered into the TargetScanHuman search box using the microRNA nomenclature: 'hsa-miR-'.

'*Human*' was selected as the species for target gene identification. A list of cumulative weighted context++ scores for each microRNA target gene was generated using TargetScanHuman. From target genes obtained for each microRNAs, only genes with a target prediction score of 90 % or greater were selected for further analysis. The individual gene lists obtained from the two platforms (i.e. miRDB and TargetScanHuman) was combined and duplications were removed in Microsoft Excel.

2.8 Annotation of micoRNA target genes

The filtered gene list was used as search parameter in DAVID https://david.ncifcrf.gov/home.jsp. The genes of interest were uploaded using "Start Analysis". "OFFICIAL_GENE_SYMBOL" was chosen as the identifier of choice and the "gene list" option was selected for viewing purposes. Candidate genes were analysed by using "Functional Annotation Clustering", from DAVID's functional annotation tools. Annotational clustering of genes was done by using "GO terms" to select annotation clusters. Since the *Cellular Component* was the GO term of interest, clusters were queried using the following terms: "membrane-bounded vesicle'', ''vesicle'', "cytoplasmic vesicle'' and "cytoplasmic membrane-bounded vesicle". This would allow for selection of gene products that can be detected within bodily fluids as they are shed.

2.9 Text mining by using Human Protein Atlas

Further analysis was done on these microRNA target genes using Human Protein Atlas. A search was done on each of the genes shortlisted by DAVID. For each gene, the words “cancer atlas” was selected, and only genes that showed high expression in BC were retained on the list. The genes resulting from the Human Protein Atlas analysis was saved as a text file.

2.10 Analysis of gene/protein interaction networks by using STRING database

Gene-protein interactions were analysed for the shortlisted genes using the STRING database as described in section 2.2.4.2. STRING was used for the analysis of gene products targeted by the selected microRNAs that are involved in BC to generate interaction networks of those proteins with known cancer proteins. Using the ‘*Multiple proteins’* option, the gene IDs for the target genes were used as input and searched for in ‘*Homo sapiens’* as the choice of organism. The STRING DB Version 10 was used for this analysis (Franceschini *et al*., 2013). Furthermore, the following criteria were used to produce each of the interaction networks: (a) a confidence level of 0.7, and (b) restricting to show no more than 20 interactions between the genes targeted by the microRNAs of interest.

2.11 Pathway analysis gene expression by using Gene Expression Atlas

Using GEA, the genes related to the microRNAs of interest were searched for in '*Homo sapiens'* as the *Organism* and both 'breast' and 'breast cancer' as the '*Sample property*' were selected. Thereafter, a filter was applied to only show genes with either 'high' or 'medium' expression.

2.12 Results

2.12.1 Data retrieval for microRNAs involved in breast cancer

Data for microRNAs which were reported to be differentially expressed in BC were retrieved using three databases – miR2disease, miRCancer and GEO. Table 2.1 illustrates the number of microRNAs extracted from the three databases. The initial microRNA lists obtained from miR2diease had 253 microRNAs, miRCancer had 291 microRNAs and GEO had 213 microRNAs. The results obtained from the three databases were combined using an excel spread sheet to remove duplicates. Out of the 757 microRNAs, 324 microRNAs were found to be duplicated among the three databases and were removed leaving a total of 433 unique microRNAs. The 433 microRNAs were rigorously checked for any experimental indication that these microRNAs has been previously studied as biomarkers for BC detection using literature mining tools (EXCON, OncomiRDB, Google scholar, Diana) Amongs the 433 microRNAs,identified 383 microRNAs were be previously validated whilst 27 microRNAs were found to be potential novel biomarkers for BC management.

Table 2.1: Extraction of novel microRNAs for breast cancer management

Database name	miR2Disease	mirCancer	Gene Expression Omnibus
Number of microRNAs identified	**253**	**291**	**213**
Combined datasets from all databases	**757**		
Number of microRNAs duplicates	**324**		
Number of microRNAs after duplicates removal	**433**		
Number of microRNAs validated	**383**		
Number of microRNAs not validated	**50**		
Validated microRNAs obtained after Literature mining (EXQUON, OncomiRDB, Google Scholar Pumed, Diana)	**23**		
Final list of non-validated or novel microRNAs	**27**		

2.12.2 MicroRNA target predictions and prioritization of target genes

Two publicly available microRNA target predictions databases, TargetScanHuman and miRDB were used as platforms to identify genes associated with the 27 microRNAs. As illustrated in Table 2.2, a total of 56,700 target genes were identified using TargetScanHuman and 9,733 by miRDB. From the total of 66,433 gene targets identified, 55,580 target genes were found to be duplicated leaving 10,853 unique target genes.

The target genes obtained from the two databases were prioritized using the following criteria: only the genes with a prediction score of 90 % or higher were used for downstream analysis; gene lists with more than 900 targets were excluded and duplications were eliminated using the Excel spread sheet. Using these criteria, a final list was generated containing 1,876 gene targets for 14 microRNAs. All the genes were represented by their official gene symbols.

Table 2.2: The number of target genes of microRNA identified by TargetScan and miRDB

Data name	miRDB	TargetScanHuman
Number of target genes identified	**9733**	**56700**
Total number of target genes from both databases	**66,433**	
Number of duplicated target genes of microRNAs	**55,80**	
Number of target genes of microRNAs after duplicates were removed	**10,853**	
Number of target genes with prediction score of 90 %	**1,876**	
Number of excluded target genes which have more than 900 targets	**8,977**	
Final list of gene targets and microRNAs	**1876 targets for 14 microRNAs**	

2.12.3 Analysis of microRNA gene targets by using functional annotation

The filtered gene list of 1,876 targets was used as input for functional annotation analysis using DAVID v6.8. A classification stringency of medium was employed in DAVID. An output of 160 genes was generated used and was categorized into three Gene Ontology (GO) groups, namely; Cellular Component (CC) Biological Process (BP) and Molecular Process (MP) as shown in Figures 2.2, 2.3 and 2.4 respectively. In terms of cellular components, target genes were found to be strongly associated (< 80 %) with cytoplasmic vesicles and vesicles; and membrane or cytoplasmic membrane bound vesicles (< 70 %). Similarly, in terms of biological processes, target genes were found to be strongly associated with intracellular protein transport (< 50 %) and vascular development (< 40 %). Finally, in terms of molecular functions, target genes were found to be associated with transcription activators (< 60 %), transcription repressors (< 50 %) and enzyme activators (< 40 %).

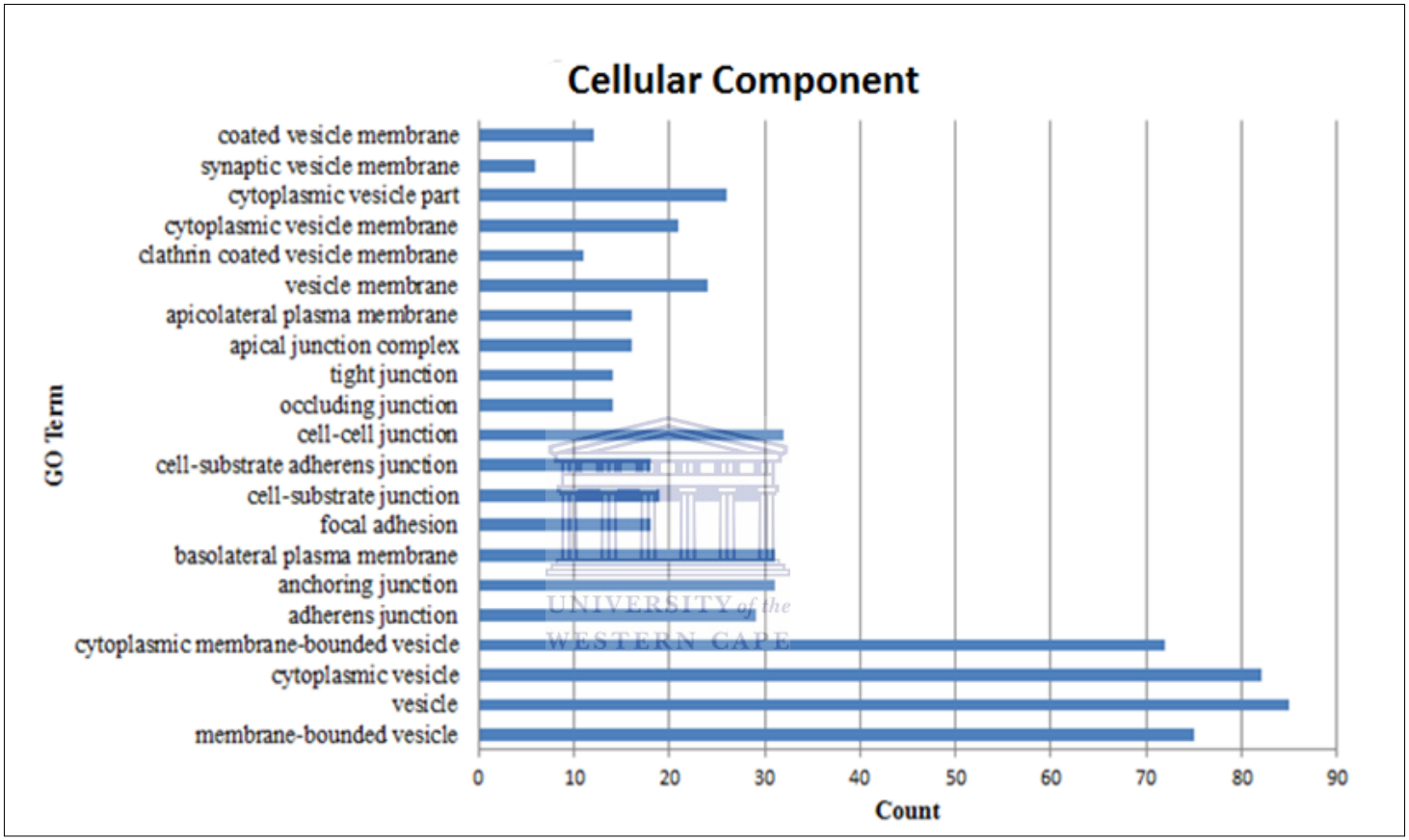


Figure 2.2: Bar graph showing the functional characterisation of microRNA target genes in DAVID based on their cellular component using GO analysis

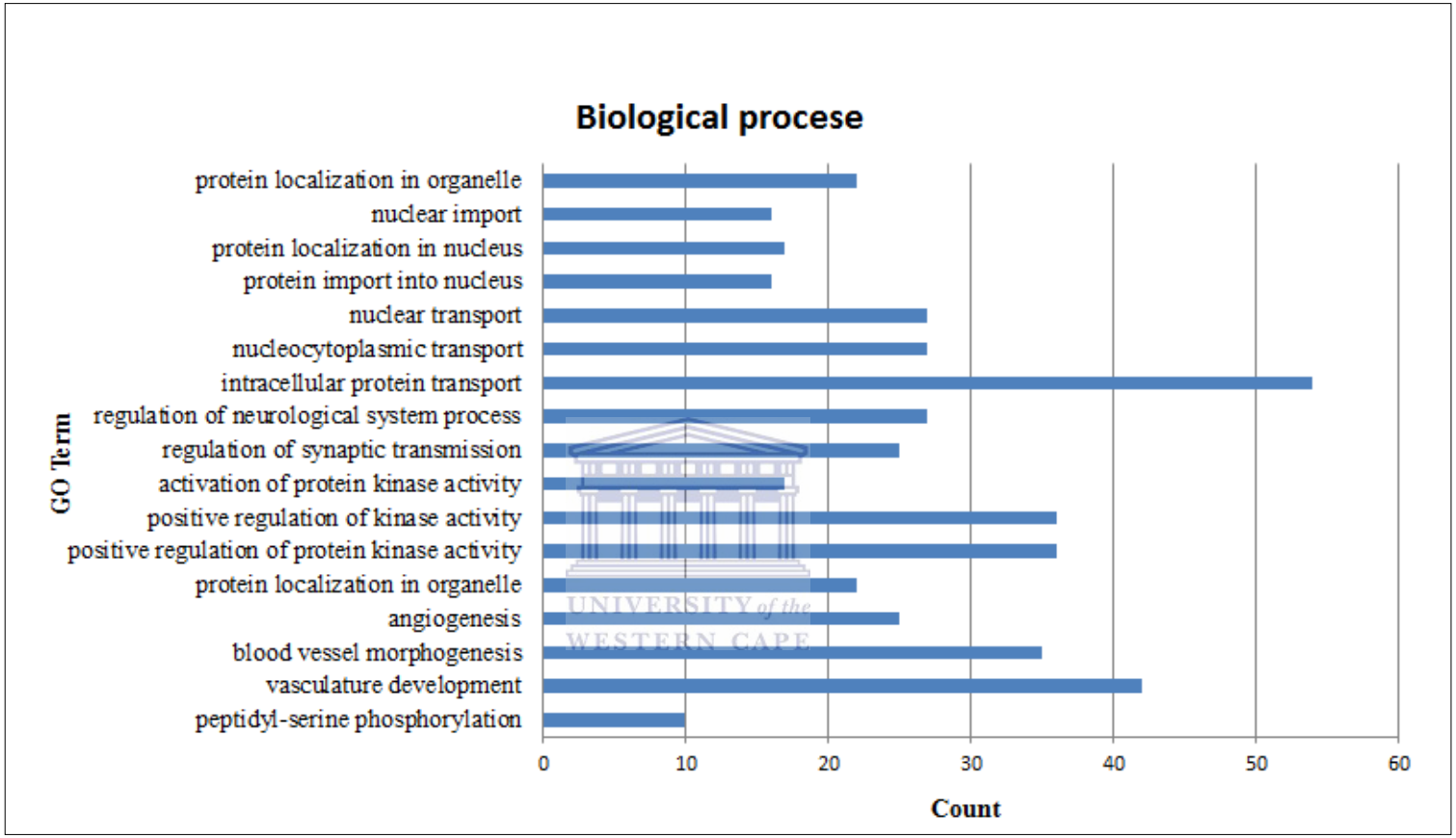


Figure 2.3: Bar graph showing the functional characterisation of microRNA target genes in DAVID based on their biological process using GO analysis.

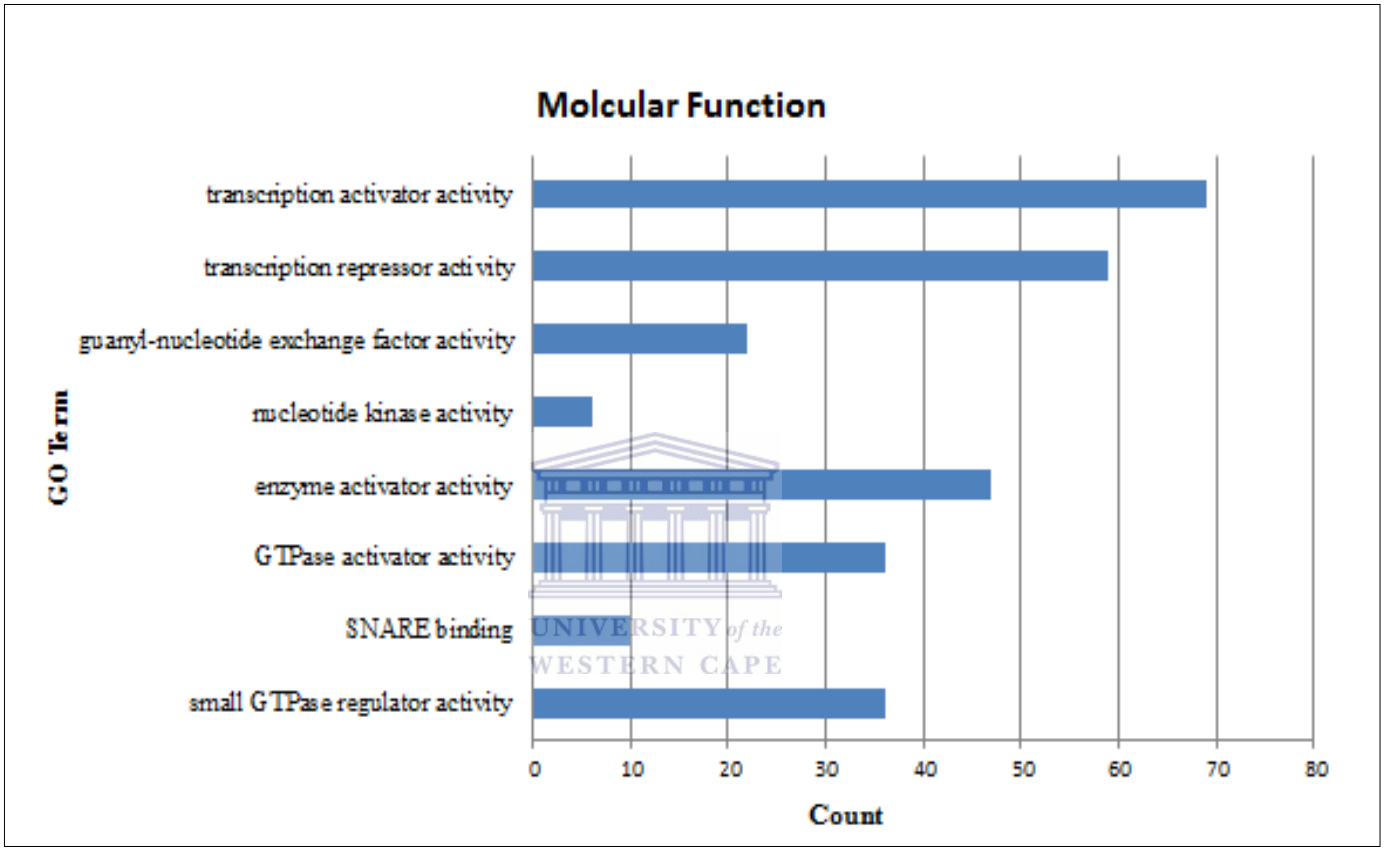


Figure 2.4: Bar graph showing the functional characterisation of genes in DAVID based on their molecular function using GO analysis.

The gene list obtained from DAVID analysis (160 genes) for 14 microRNAs were further validated for their novelty as BC biomarkers using Human Protein Atlas database as described in section 2.9. A total of 9 genes (Figure 2.5) were identified for 5 microRNAs using this analysis and this included genes that were known and predicated to be either up or down regulated/expressed in BC tissue when compared to normal breast tissue.

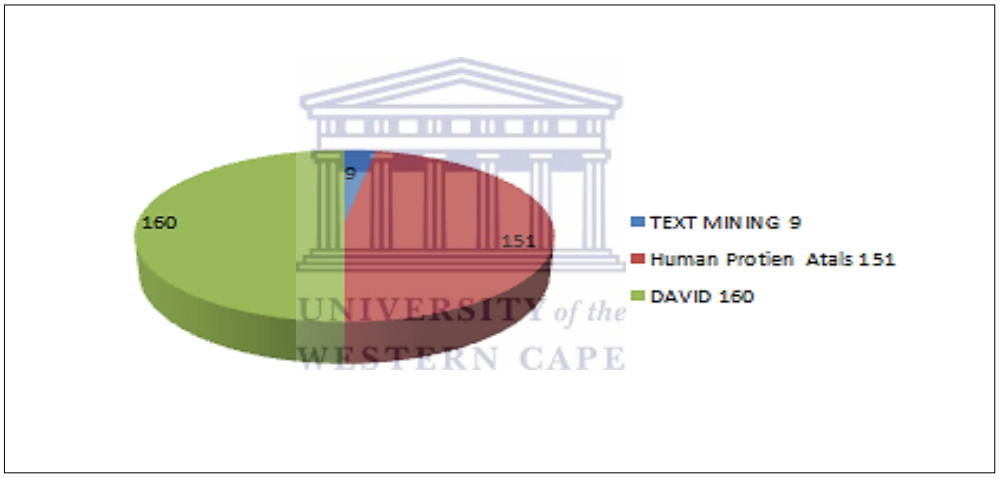


Figure 2.5: Number of target genes for microRNAs obtained from functional characterization using DAVID and tissue expression analysis using Human Protein Atlas.

2.12.4 Protein /Protein Interaction Analysis via STRING

Unique gene identifiers (*ACTN4*, *TMED10*, *STY9*, *GOPC*, *VEGFA*, *CD2AP*, *ANP32F*, *HIP1* and *AP1G1*) were used to produce expression networks using the

STRING database as described in section 2.10. The results in Figure 2.6 shows a clear link between 3 of the 9 target genes as shown in the black Box present in the overall gene network generated by using the criteria as described in section 2.10. Gene targets *VEGFA*, *CD2AP* and *ACTN4* are seen to be connected based on text mining and others databases with the remaining 6 genes not showing any interaction to one another (Figure 2.7).

In addition, cellular component analysis and biological processes were also performed on the 9 microRNA target genes using STRING. In Table 2.3, is shown that the HIP1 target gene does not have cellular component analysis and in Table 2.4 the biological process for *HIP1* and *ANP32E* is not shown (Table 2.3 and Table 2.4).

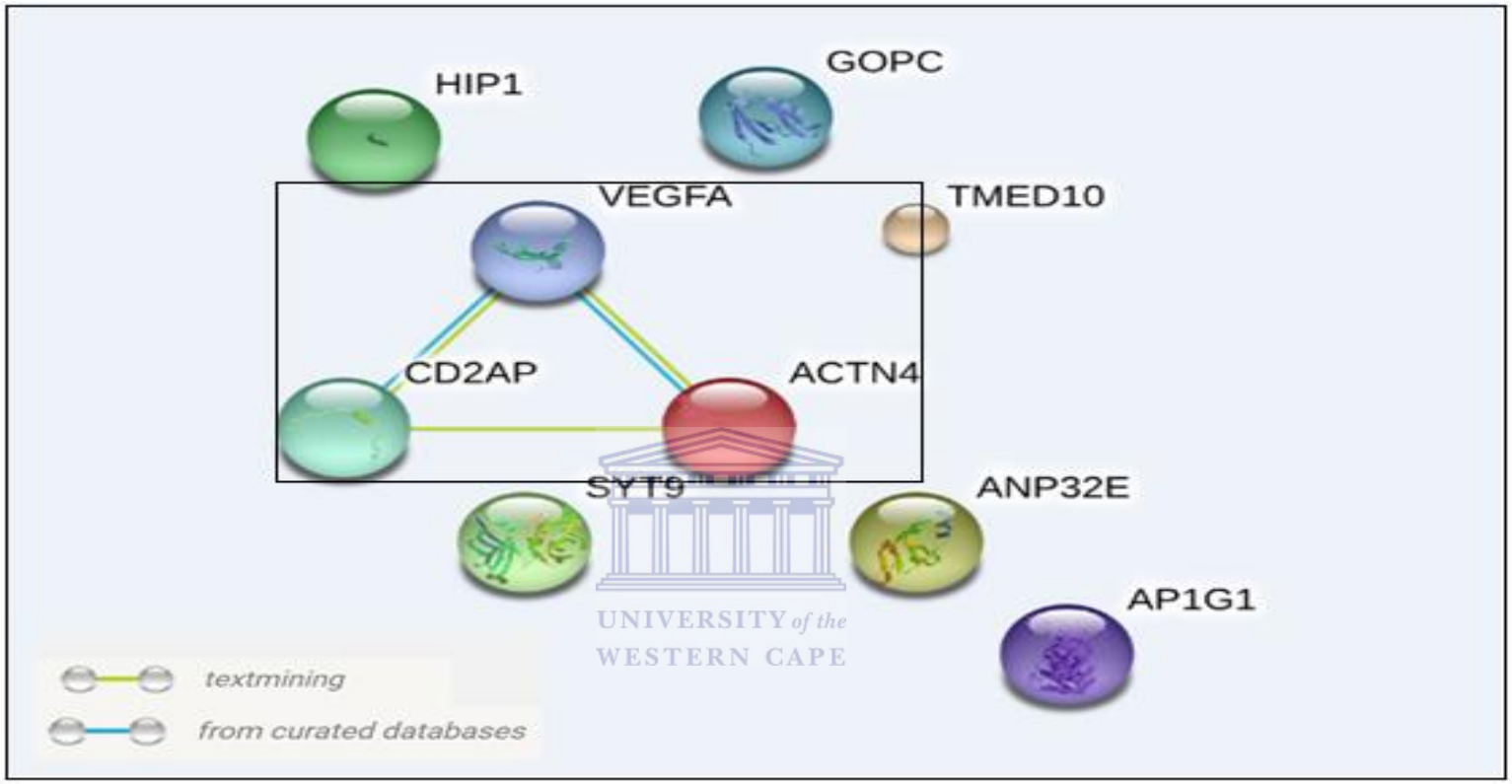


Figure 2.6: Map of the Protein Network generated by STRING.

Genes are represented by the nodes and the different colours lines represent the types of evidence for the association.

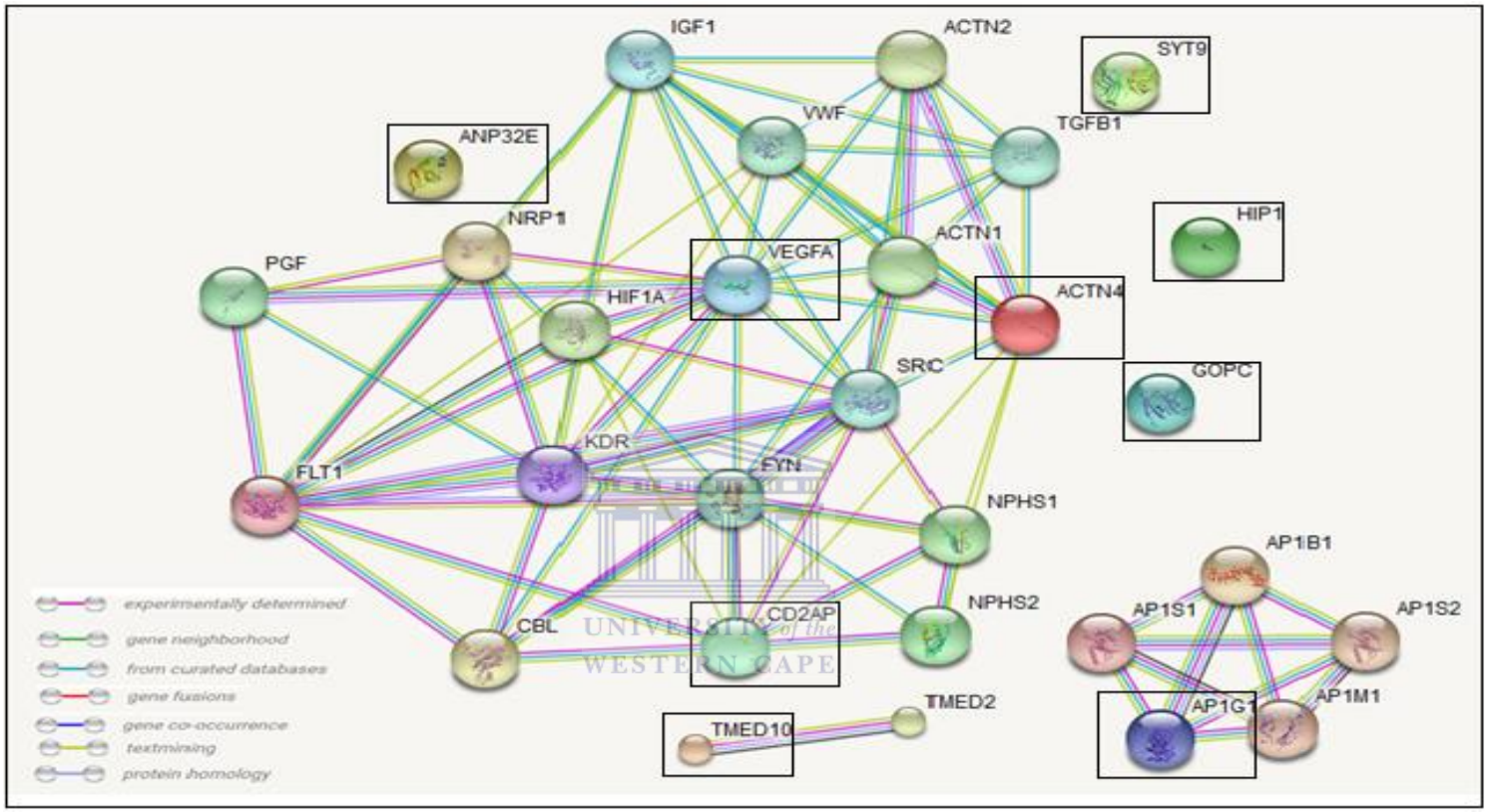


Figure 2.7: Map of the Protein interactions of the 9 microRNA target genes in association with other genes as generated by STRING. The 9 microRNA target genes are outlined in boxes.

Table 2.3: microRNA target genes and their cellular component analysed using STRING

Gene	Cellular component
ACTN4	cytoplasmic vesicle membrane ,cytoplasmic membrane-bounded vesicle, secretory granule, membrane-bounded vesicle, platelet alpha granule
GOPC	clathrin-coated vesicle ,cytoplasmic membrane-bounded vesicle, cytoplasmic vesicle part, trans-Golgi network transport vesicle, membrane-bounded vesicle
CD2AP	cytoplasmic membrane-bounded vesicle
ANP32E	Cytoplasmic membrane-bounded vesicle, membrane-bounded vesicle.
VEGFA	secretory granule ,membrane-bounded vesicle, platelet alpha granule
AP1G1	clathrin-coated vesicle, cytoplasmic vesicle membrane, cytoplasmic membrane-bounded vesicle, cytoplasmic vesicle part, membrane-bounded vesicle
TMED10	clathrin-coated vesicle, cytoplasmic vesicle membrane, cytoplasmic membrane-bounded vesicle, secretory granule, trans-Golgi network transport vesicle, membrane-bounded vesicle
SYT9	secretory granule membrane, cytoplasmic vesicle membrane, cytoplasmic membrane-bounded vesicle secretory granule, membrane-bounded vesicle

Table 2.4: microRNA target genes involved in various biological processes analysed using STRING.

Gene	Biology process
ACTN4	regulation of vesicle-mediated transport, exocytosis, vesicle-mediated transport
GOPC	regulation of vesicle-mediated transport, exocytosis
CD2AP	regulation of vesicle-mediated transport
VEGFA	regulation of vesicle-mediated transport, exocytosis, vesicle-mediated transport,
AP1G1	regulation of vesicle-mediated transport ,exocytosis
TMED10	Exocytosis
SYT9	exocytosis, vesicle-mediated transport

2.12.5 Pathway analysis of the genes via Gene Expression Atlas

The GEA database was used to identify differential expression of the nine identified genes in BC. Differential expression of all the nine genes were studied as described in section 2.11 comparing gene expression in BC tissues to that in normal tissues. Figures 2.8 – 2.16 display the expression analysis of *ACTN4*, *TMED10*, *STY9*, *GOPC*, *VEGFA*, CD2AP, *ANP32F*, *HIP1* and *AP1G1 respectively*. A medium expression level was found for *ACTN4*, *TMED10*, *STY9*, *GOPC* and *VEGFA* both in normal breast tissue and cancerous breast tissue. On the other hand *CD2AP*, *ANP32F*, *HIP1* and *AP1G1* showed medium expression in BC tissue and low expression in normal breast tissue.

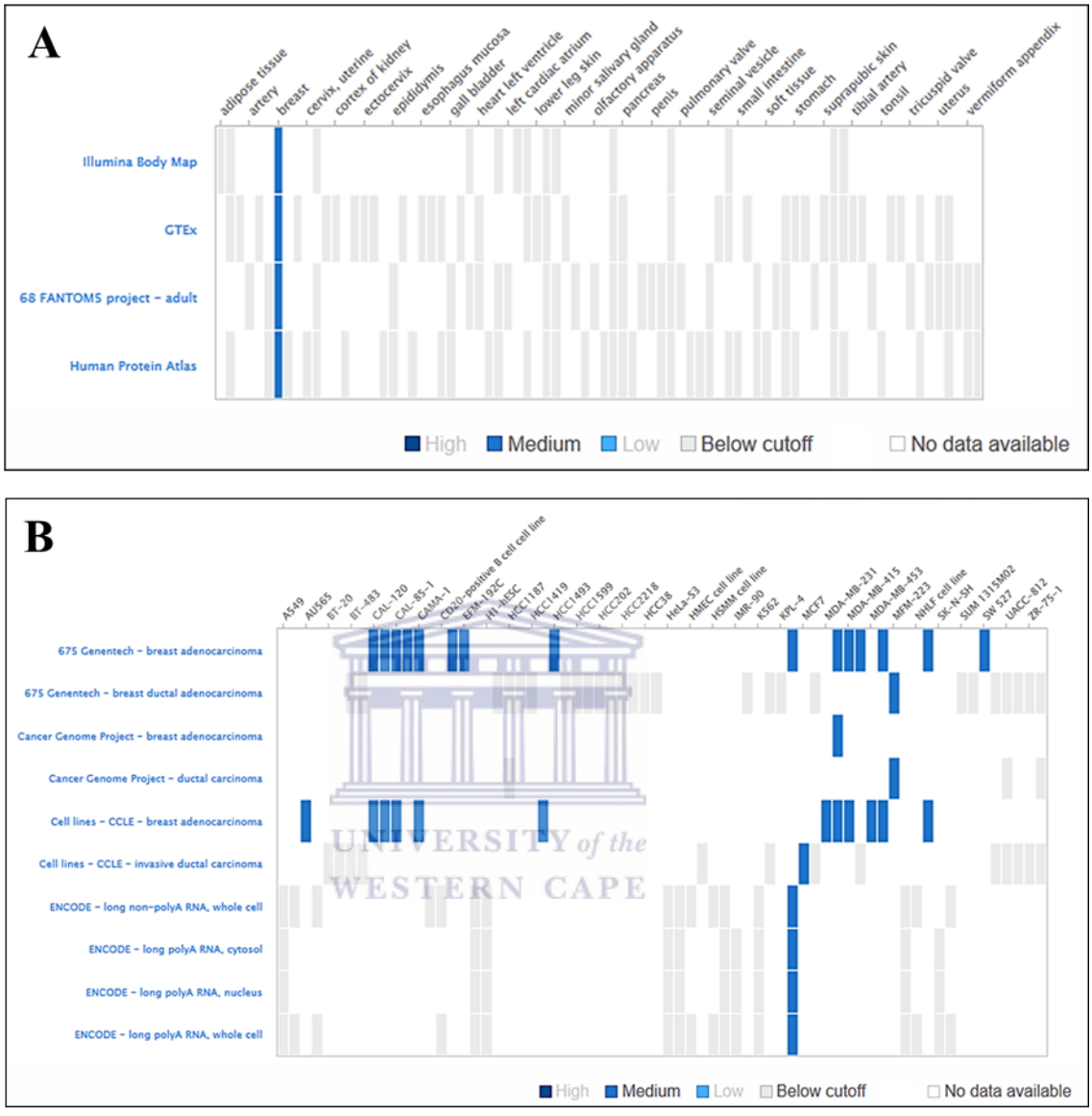


Figure 2.8: Expression profile for *ACTN4* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancer tissue.

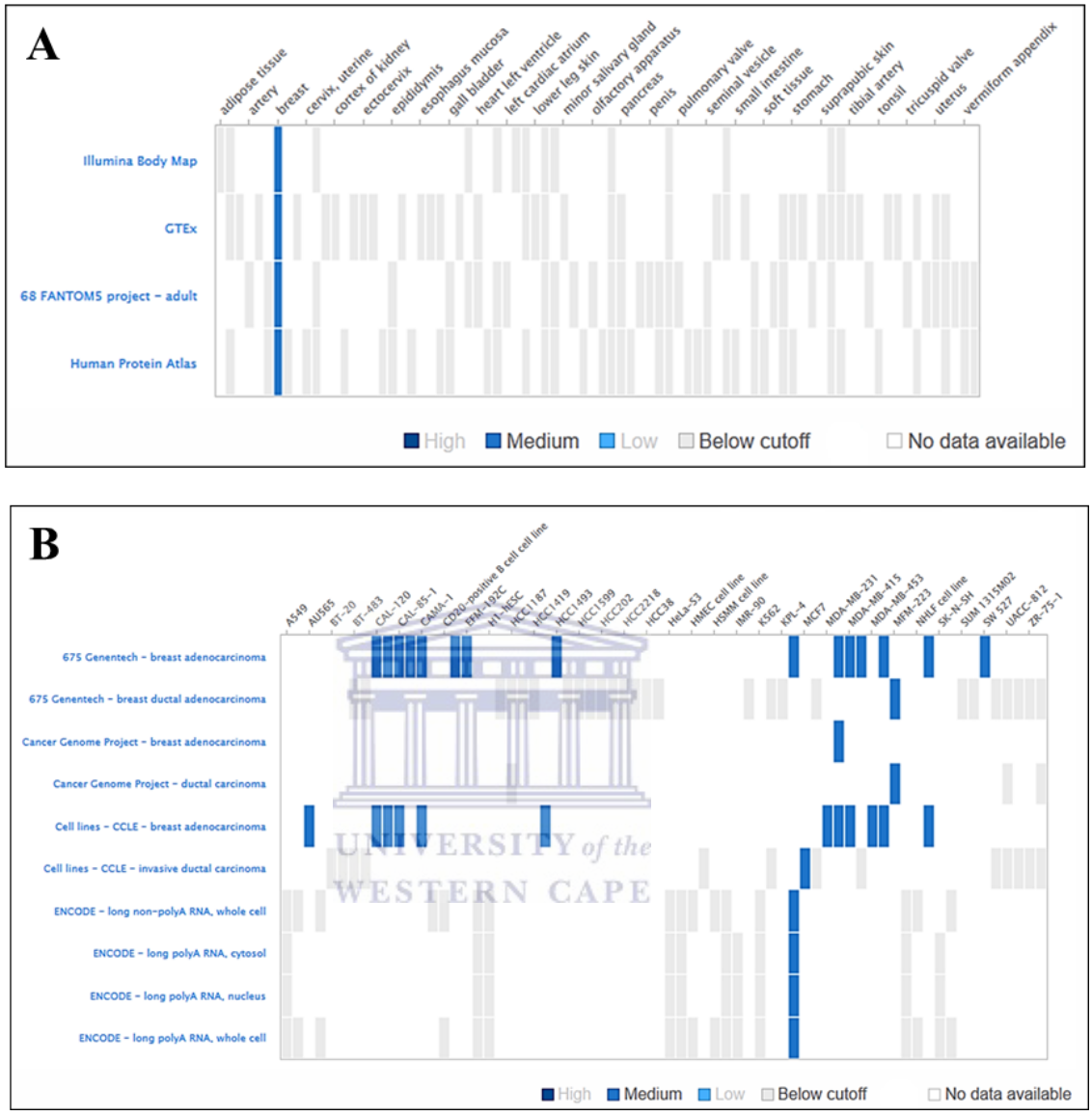


Figure 2.9: Expression profile for *AP1G1* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancer tissue.

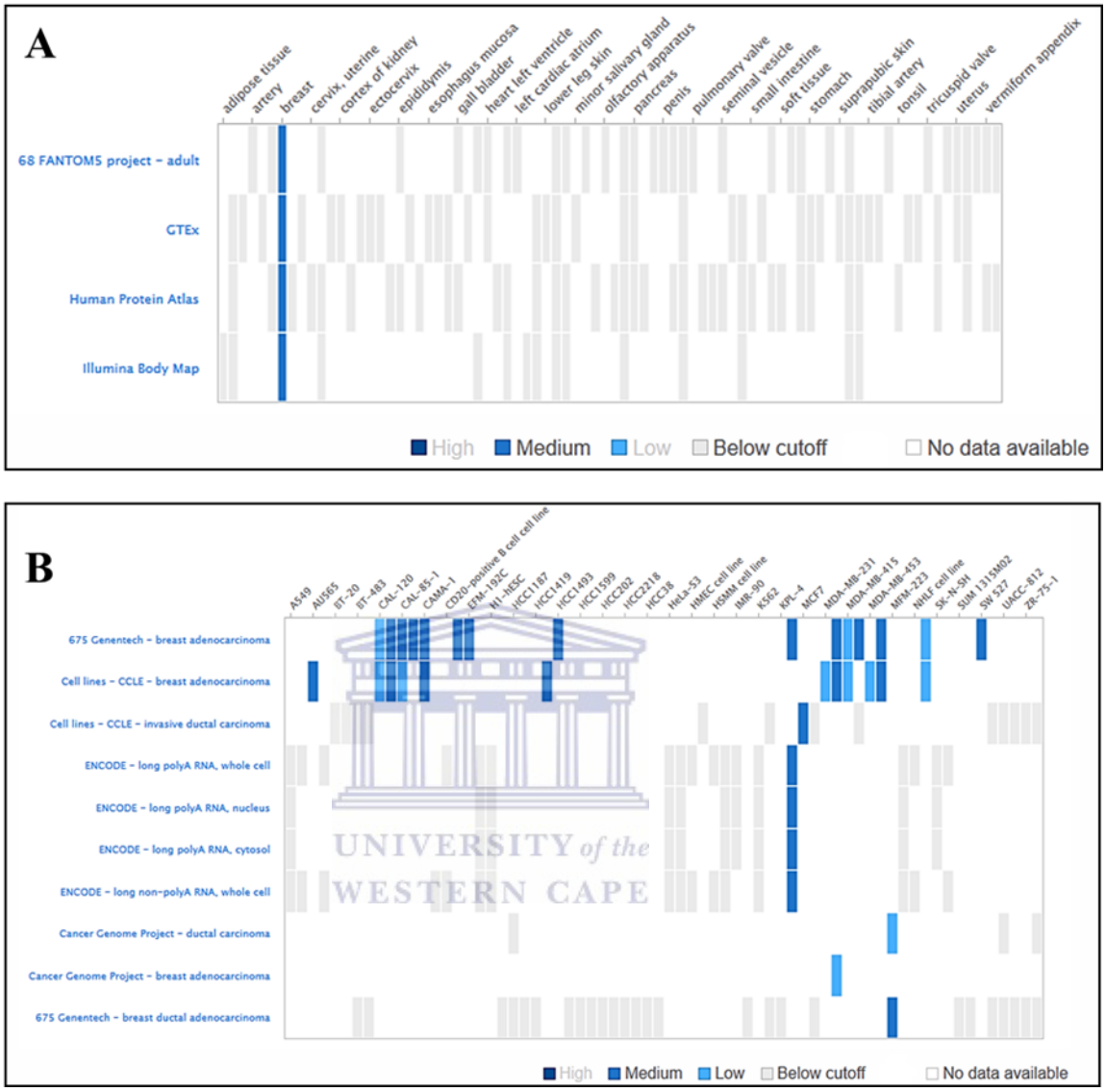
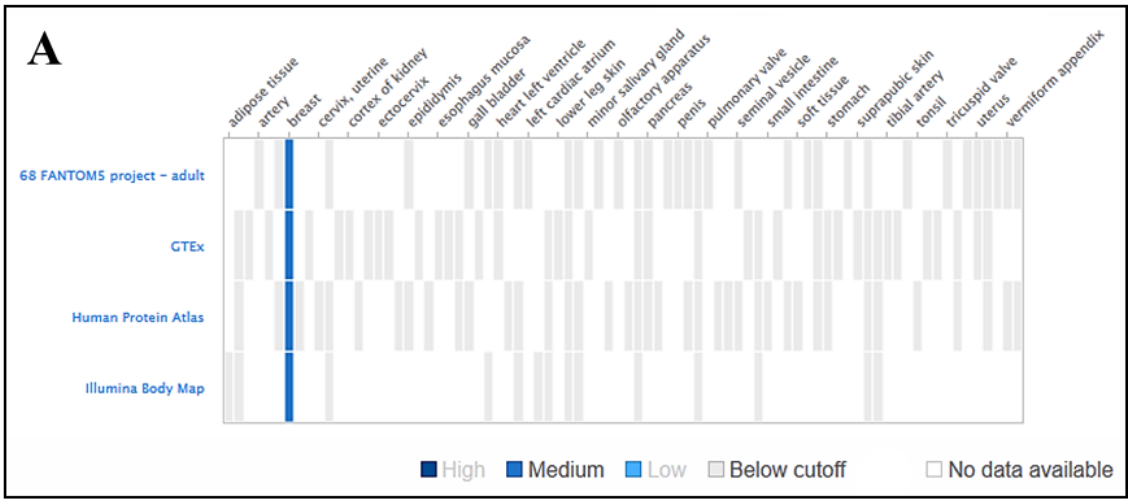


Figure 2.10: Expression profile for *HIP1* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancer tissue.



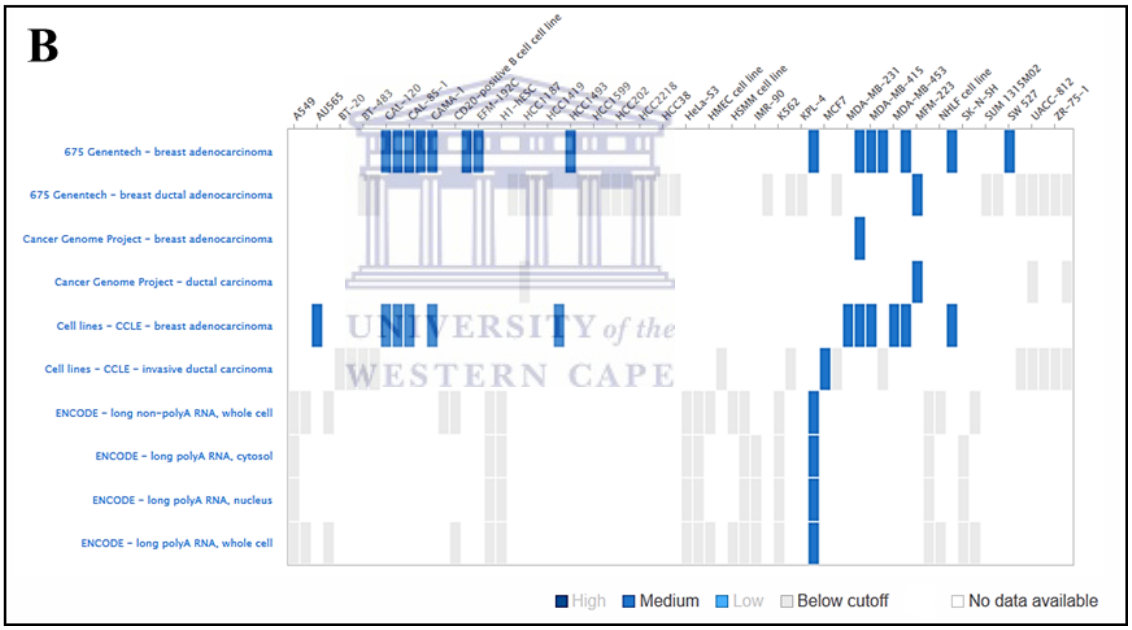


Figure 2.11: Expression profile for *TMED10* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancertissue.

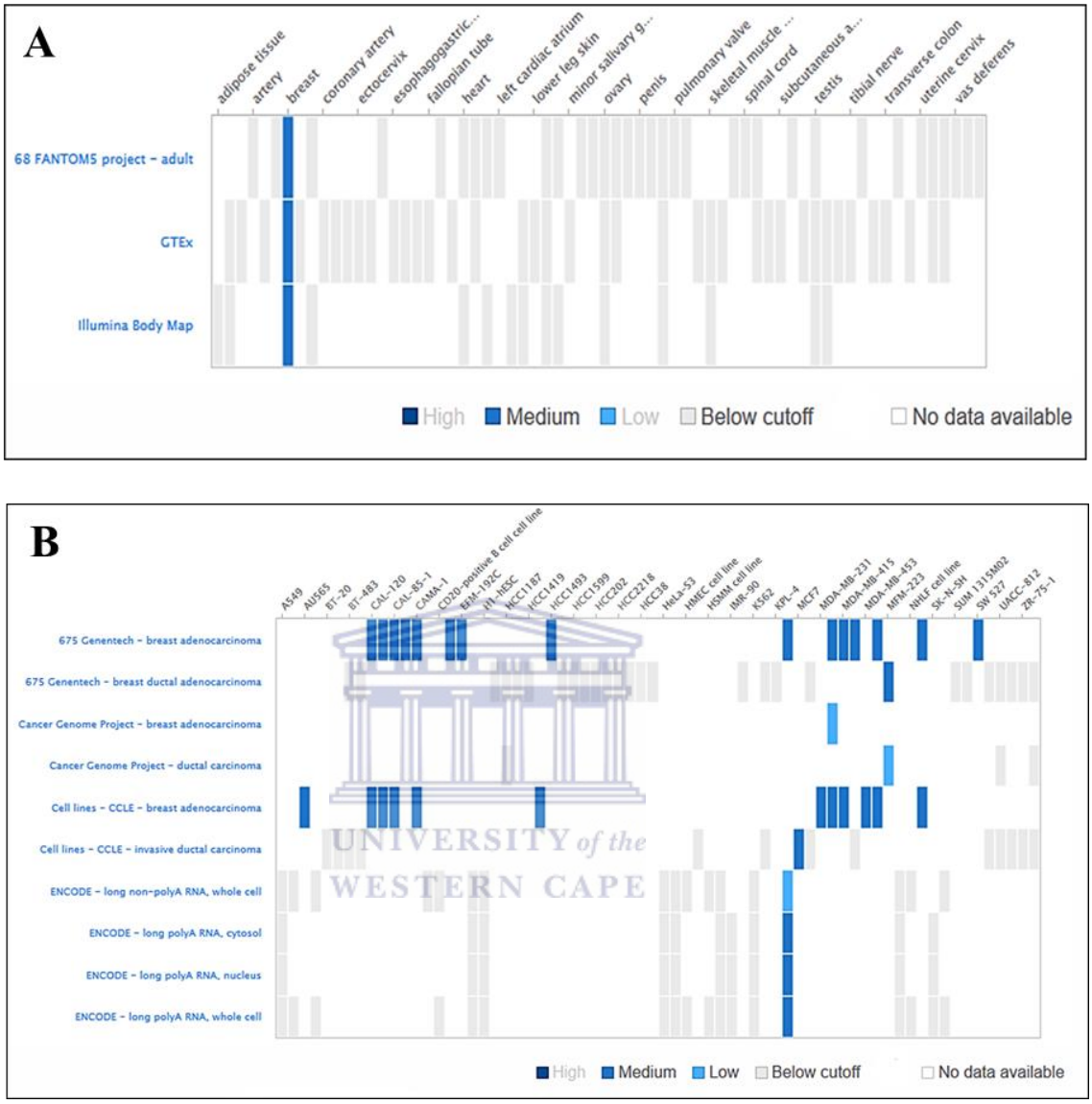


Figure 2.12: Expression profile for *GOPC* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancer tissue.

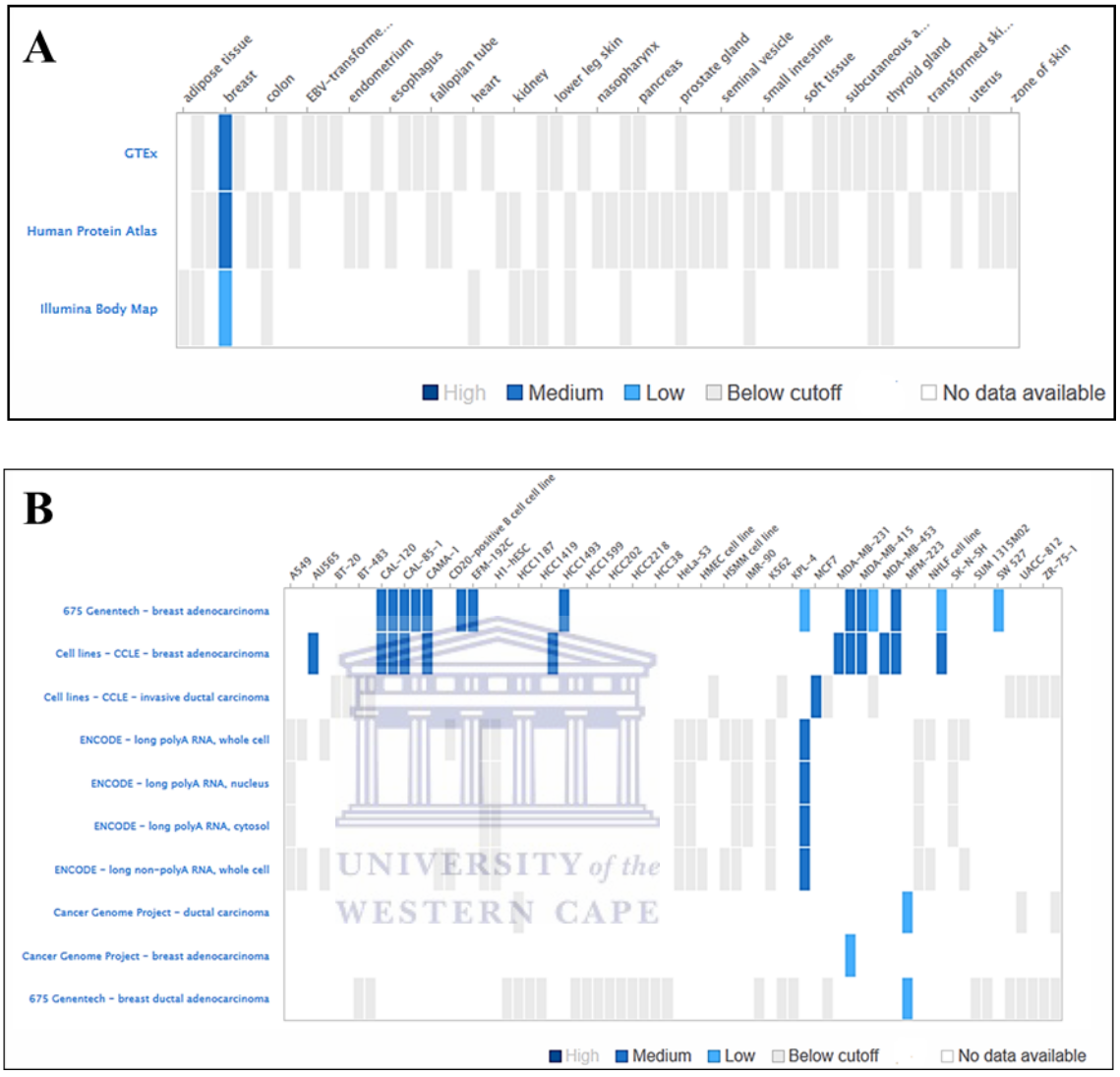


Figure 2.13: Expression profile for *VEGF* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) breast cancer tissue.

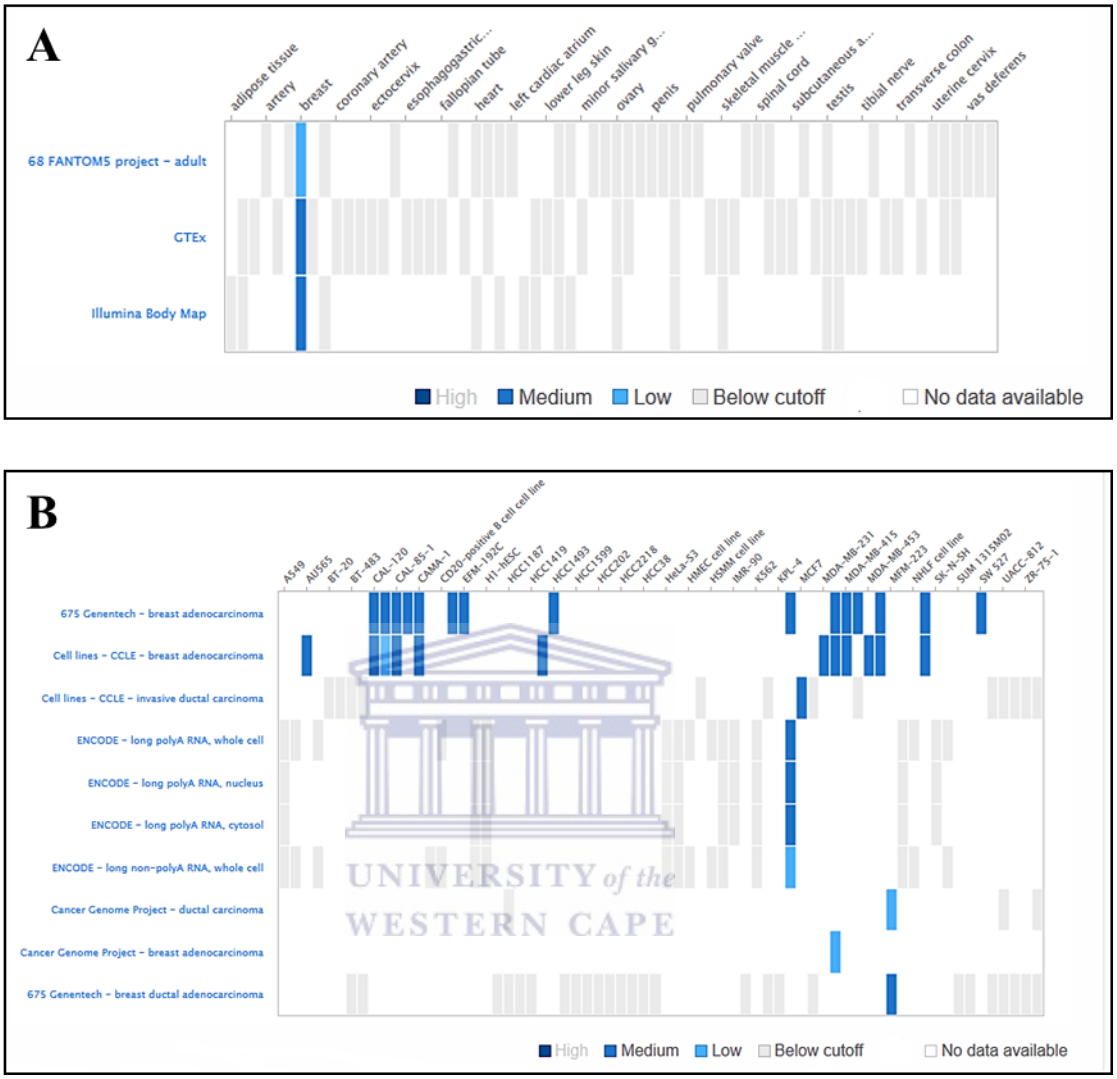


Figure 2.14: Expression profile for *ANP32E* from Gene Expression Atlas (GEA) (A) normal breast and (B) in breast cancer tissue.

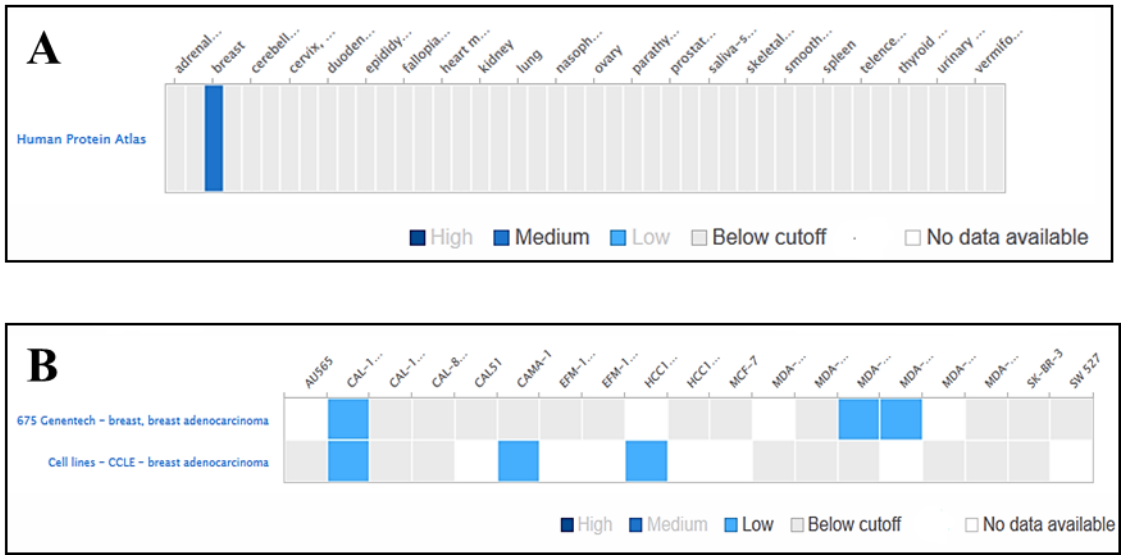


Figure 2.15: Expression profile for *STY9* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) in breast cancer tissue.

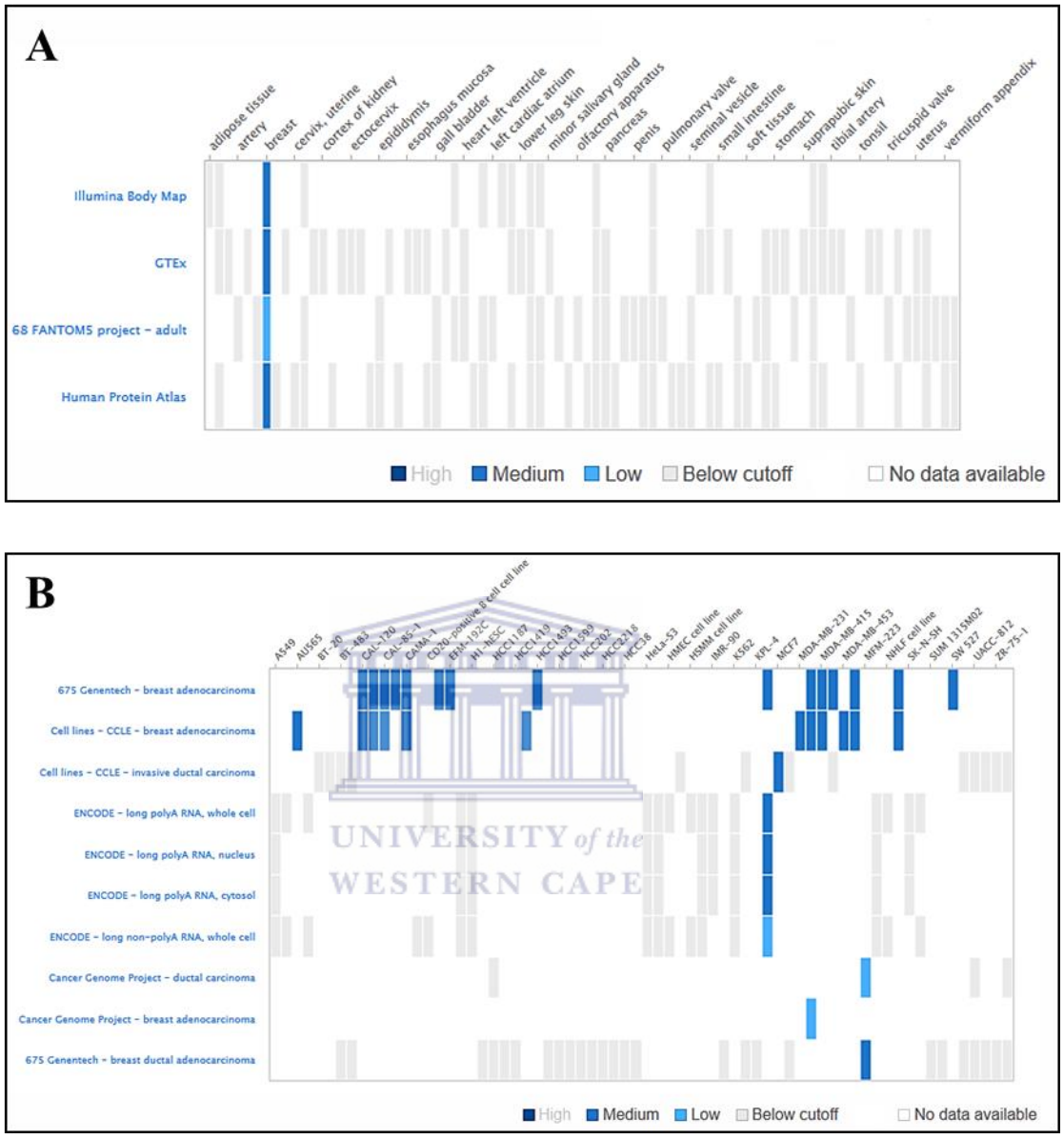


Figure 2.16: Expression profile for *CD2AP* from Gene Expression Atlas (GEA) (A) normal breast tissue and (B) in breast cancer tissue.

Table 2.5: MicroRNAs and their specific targets gene

MicroRNA	Target Genes
MiR1	***HIP1***
MiR2	***SYT9***
MiR3	***TMED10, AP1G1***
MiR4	***VEGFA, ANP32E, CD2AP***
MiR5	***ACTN4, GOPC***

2.13 Discussion

Breast cancer (BC) has been associated with death amongst women worldwide including South Africa. The etiology of BC is highly heterogeneous in nature and involves aberrations in a network of pathways associated with fundamental mechanisms like proliferation, differentiation, migration, maturation and apoptosis. Almost all of these key pathways are found to be critically regulated by microRNAs (miRNAs), which are non-coding RNAs and post-transcriptional (Bertoli *et al*., 2013). These miRNAs were shown to be down-regulated in a large number of tumours including BC. Identification of specific microRNAs along with their physiological functions provides a promising approach in cancer diagnosis, prognosis and treatment. In particular, automated comparison of the gene expression profiles of

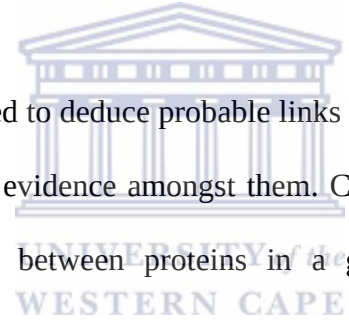
diseased versues. normal state of cancer can be used to determine differentially expressed genes particulate in BC which can be further implicated in networks related to BC to identify the disease mechanisms. In the last decades, advancements in bioinformatics tools has supported the creation of huge repository of human transcriptomics data which facilitated the quick and logical investigations of biological and disease mechanisms, including BC. The present study was aimed at identifying novel microRNAs that can potentially be used for the early detection of BC, prognosis and management of the disease.

Several *in silico* tools were employed to identify novel microRNAs that can potentially be implicated in BC and its mechanistic pathways. A list of microRNAs was produced by selecting microRNAs not molecularly validated and associated with BC using the following databases: miR2Disease, miRCancer and Gene Expression Omnibus. The target genes of these shortlisted microRNAs were identified using TargetScan and miRDB. The algorithm in TargetScan Human 6.2 scores its microRNA target gene based on the level of conserved 8-mer and 7-mer sites that match the seed region of each microRNA (Li *et al*., 2014). Furthermore, the specificity of this software for the target site is also dependant on its determinants, such as type site contribution, which calculates the score for both 7- and 8-mer motifs. 8-mer motifs are allotted a higher score as they are more down-regulated than those with 7-mer motifs (Radfar, 2014). It should be noted that a more negative context score is associated with a more favourable binding target site (Garcia *et al*.,

2011). Additionally, TargetScan also ranks the target genes according to their probability of conserved targeting (PCT). PCT values with a lower probabilistic value than 52 would decrease the integrity of conservation for the predicted binding target sites across multiple species (Friedman *et al*., 2009). The use of two prediction algorithms for microRNAs target genes identification reduces the false positive target predictions and enhances the sensitivity and accuracy of presently used methodology as observed from this study results where only 14 microRNAs and 1,876 target genes were selected for further studies from an initial list of 757 microRNAs and 66,433 target genes.

Functional annotation analysis using DAVID 6.8 showed that the 1,876 targets of 14 microRNA could be implicated in cancer pathways through Gene Ontology (GO) analysis. In terms of cellular components, target genes were found to be strongly associated with cytoplasmic vesicles and vesicles; and membrane or cytoplasmic membrane bound vesicles. Tumour-derived microvesicles have been implicated in disease progression. It is reported that heterogeneous membrane-bound sacs are shed as well as taken up by tumour cells. While the dispersed microvesicles are thought to be responsible for disease progression, its uptake helps in the maintenance of the tumour microenvironment (D'Souza-Schorey & Clancy, 2012). Similarly, in terms of biological processes, target genes were found to be strongly associated with intracellular protein transport and vascular development. Dysregulation of nucleo-cytoplasmic transport, particularly tumour suppressor and oncoproteins has been

reported in cancer (Hill *et al*., 2014). Finally, in terms of molecular functions, target genes were found to be associated with transcription activators, transcription repressors and enzyme activators. Genome instability due to mutations in transcription factors and noncoding RNAs is one of the hallmarks of almost all cancers, including BC. Aberrations in transcription activation, repression and regulation can contribute to tumourigenesis with overall alteration of core auto-regulatory circuitry of the cell (Lee & Young, 2013). Overall, all these mechanisms have great potential in early diagnosis of BC.



STRING analysis was used to deduce probable links between the nine finalized genes based on several lines of evidence amongst them. Complete knowledge of all direct and indirect interactions between proteins in a given cell would represent an important milestone towards a comprehensive description of cellular mechanisms and functions (Franceschini *et al*., 2013). A clear link between three of the nine target genes (*VEGFA*, *CD2AP* and *ACTN4*) were obtained with no connections found between the remaining 6 genes (see Figure 2.6). The lack of connections may perhaps be due to the fact that STRING database either have partial information about the genes or may not contain enough experimental information to confirm the link between these genes. In Tables 2.3 and 2.4 the target genes (HIP1 and ANP32E) are not shown The reason for this might be because the database was in the process of being updated or that genes have not enough information that is linked to cancer and specifically BC. Recent studies of *ACTN4* demonstrated not only a role in cancer

invasion, but also its biological role as a transcriptional co-activator. *ACTN4* is aggressively involved in the tumourigenesis of BC, and this concept is attracting a lot of attention. The localization of actinin-4 in the nucleus is very interesting in terms of tumourigenesis (Honda, 2015). Furthermore, in the present study *VEGFA* was found to be the most prominent interacting protein in the derived network (see Figure 2.7). Vascular endothelial growth factor-A (*VEGF-A*) are reported to play crucial roles in tumour angiogenesis and cancer progression. In a study conducted by Balasubramanian et. al., 2007 variations in *VEGF* gene expression have been reported to be involved in BC susceptibility and severity (Balasubramanian *et. al*., 2007). Finally, CD2-associated protein (*CD2AP*) has been reported to be involved in cytoskeletal remodelling (Weber *et al*., 2004; Shih *et al.,* 1999), cell survival (Huber *et. al.,* 2003; Schiffer *et al*., 2004), and endocytosis (Cormont *et al.,* 2003; Lynch *et al*., 2003; Kobayashi *et al*., 2004). *CD2AP* is a 80-kDa protein, which interacts with the cytoplasmic domain of $CD2^+$ T lymphocyte and natural killer-specific membrane protein (Dustin *et. al*., 1998). Based on several studied $CD2^+$ T lymphocytes have been implicated in bladder cancer and cervical cancer (Sandquist *et al*., 2015; Ting *et al*., 2010). Overall, these results clearly demonstrates the role of these genes in BC progression and hence identification of their regulating microRNAs can greatly help in BC diagnosis.

In addition, using the GEA database, differential expression of *ACTN4*, *TMED10*, *STY9, GOPC*, *VEGFA*, *CD2AP*, *ANP32F*, *HIP1* and *AP1G1* genes were studied both in normal breast tissue and BC tissue. Medium expression level was observed for

genes *ACTN4, TMED10, STY9, GOPC* and *VEGFA* both in normal breast tissue in BC. On the other hand *CD2AP*, *ANP32F*, *HIP1* and *AP1G1* were found to show medium expression in BC tissue and low expression in normal breast tissue.

To conclude, the present study findings derived through combination of functional annotation and expression profiling tools resulted in the identification of nine target genes regulated by five microRNAs which may play key roles in BC development as well as progression. Further validation of the identified microRNAs as biomarkers will be performed in subsequent chapters.

2.14 References

Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, 4, p.e05005.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. and Harris, M.A., 2000. Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), pp.25-29.

Balasubramanian, S.P., Cox, A., Cross, S.S., Higham, S.E., Brown, N.J. and Reed, M.W., 2007. Influence of VEGF-A gene variation and protein levels in breast cancer susceptibility and severity. *International journal of cancer*, 121(5), pp.1009-1016.

Barrett, T. and Edgar, R., 2006. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. *Methods in enzymology*, 411, pp.352-369.

Bertoli, G., Cava, C. and Castiglioni, I., 2015. MicroRNAs: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. *Theranostics*, 5(10), p.1122.

Cormont, M., Metón, I., Mari, M., Monzo, P., Keslair, F., Gaskin, C., McGraw, T.E. and Marchand-Brustel, L., 2003. CD2AP/CMS regulates endosome morphology and traffic to the degradative pathway through its interaction with Rab4 and c-Cbl. *Traffic*, 4(2), pp.97-112.

Dennis, 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 biology*, 4(9), p.R60.

D'Souza-Schorey, C. and Clancy, J.W., 2012. Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. *Genes & development*, 26(12), pp.1287-1299.

Dustin, M.L., Olszowy, M.W., Holdorf, A.D., Li, J., Bromley, S., Desai, N., Widder, P., Rosenberger, F., van der Merwe, P.A., Allen, P.M. and Shaw, A.S., 1998. A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. *Cell*, 94(5), pp.667-677.

Exiqon miRSearch Help Page. Accessed at http://www.exiqon.com/ls/Pages/TPTHelpPage.aspx 19/2/2017.

Fendler, A., 2011 MiRNAs and prostate cancer*: identification, functional characterization and their potential use in medical practice* (Doctoral dissertation, Berlin, Freie Universität Berlin, Diss., 2011).

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), pp.D808-D815..

Friedman, R.C., Farh, K.K.H., Burge, C.B. and Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, 19(1), pp.92-105.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A. and Bartel, D.P., 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nature structural & molecular biology*, 18(10), pp.1139-1146.

Grimson, A., Farh, K.K.H., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, 27(1), pp.91-105.

Hill, R., Cautain, B., de Pedro, N. and Link, W., 2014. Targeting nucleocytoplasmic transport in cancer therapy. *Oncotarget*, 5(1), pp.11-28.

Honda, K., 2015. The biological role of actinin-4 (ACTN4) in malignant phenotypes of cancer. *Cell & bioscience*, 5(1), p.41.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, 8(9), p.R183.

Huber, T.B., Hartleben, B., Kim, J., Schmidts, M., Schermer, B., Keil, A., Egger, L., Lecha, R.L., Borner, C., Pavenstädt, H. and Shaw, A.S., 2003. Nephrin and CD2AP

associate with phosphoinositide 3-OH kinase and stimulate AKT-dependent signaling. *Molecular and cellular biology*, 23(14), pp.4917-4928.

James,.(2013). "Google Scholar Library". Google Scholar Blog. Accessed at https://scholar.googleblog.com/2013/11/google-scholar-library.html on 3/1/2017

Ji Diana Lee, Yun, Veronica Kim, Dillon C. Muth, and Kenneth W. Witwer. "Validated microRNA target databases: an evaluation." *Drug development research* 76, no. 7 (2015): 389-396.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. and Liu, Y., 2009. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic acids research*, 37(suppl 1), pp.D98-D104.

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*, p.gkp936.

Kobayashi, S., Sawano, A., Nojima, Y., Shibuya, M. and Maru, Y., 2004. The c-Cbl/CD2AP complex regulates VEGF-induced endocytosis and degradation of Flt-1 (VEGFR-1). *The FASEB Journal*, 18(7), pp.929-931.

Lee, T.I. and Young, R.A., 2013. Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), pp.1237-1251.

Lekprasert, P., 2012. *MicroRNA target prediction via duplex formation features and direct binding evidence* (Doctoral dissertation, Duke University).

Lewis, B.P., Burge, C.B. and Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *cell*, 120(1), pp.15-20.

Lynch, D.K., Winata, S.C., Lyons, R.J., Hughes, W.E., Lehrbach, G.M., Wasinger, V., Corthals, G., Cordwell, S. and Daly, R.J., 2003. A Cortactin-CD2-associated protein (CD2AP) complex provides a novel link between epidermal growth factor receptor endocytosis and the actin cytoskeleton. *Journal of Biological Chemistry*, 278(24), pp.21805-21813.

Persson, A., Hober, S. and Uhlen, M., 2006. A human protein atlas based on antibody proteomics. *Current opinion in molecular therapeutics*, 8(3), pp.185-190.

Pontén, F., Jirström, K. and Uhlen, M., 2008. The Human Protein Atlas—a tool for pathology. *The Journal of pathology*, 216(4), pp.387-393.

Pontén, F., Schwenk, J.M., Asplund, A. and Edqvist, P.H., 2011. The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal of internal medicine*, 270(5), pp.428-446.

Radfar, M.H., 2014. *Computational prediction of target genes of microRNAs* (Doctoral dissertation, University of Toronto).

Raza, K., 2012. Application of data mining in bioinformatics. *arXiv preprint arXiv:1205.1125*.

Sandquist, E., Somji, S., Sens, D. and Garrett, S., 2015. N-cadherin Upregulation and EMT Progression in As3+ and Cd2+-Transformed Urothelial Cells. *The FASEB Journal*, 29(1 Supplement), pp.629-8.

Schiffer, M., Mundel, P., Shaw, A.S. and Böttinger, E.P., 2004. A novel role for the adaptor molecule CD2-associated protein in transforming growth factor-β-induced apoptosis. *Journal of Biological Chemistry*, 279(35), pp.37004-37012.

Schmitz, U., Lai, X., Winter, F., Wolkenhauer, O., Vera, J. and Gupta, S.K., 2014. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic acids research*, 42(12), pp.7539-7552.

Shih, N.Y., Li, J., Karpitskii, V., Nguyen, A., Dustin, M.L., Kanagawa, O., Miner, J.H. and Shaw, A.S., 1999. Congenital nephrotic syndrome in mice lacking CD2-associated protein. *Science*, 286(5438), pp.312-315.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. and Kuhn, M., 2014. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, p.gku1003.

Ting, Z., Fei, X. and Li, Z., 2010. Clinical Significance of Determination of Peripheral Blood Mononuclear Cells CD2~+ Expression in Patients with Uterine Cervix Cancer. *Journal of Radioimmunology*, *5*, p.043.

Uhlen, M. and Ponten, F., 2005. Antibody-based proteomics for human tissue profiling. *Molecular & Cellular Proteomics*, 4(4), pp.384-393.

Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.L., Maniou, S., Karathanou, K., Kalfakakou, D. and Fevgas, A., 2015. DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. *Nucleic acids research*, 43(D1), pp.D153-D159.

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), pp.D433-D437.

Wang, D., Gu, J., Wang, T. and Ding, Z., 2014. OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics*, 30(15), pp.2237-2238.

Weber, S., Gribouval, O., Esquivel, E.L., Morinière, V., Tête, M.J., Legendre, C., Niaudet, P. and Antignac, C., 2004. NPHS2 mutation analysis shows genetic heterogeneityof steroid-resistant nephrotic syndrome and lowpost-transplant recurrence. *Kidney international*, 66(2), pp.571-579.

Wong N and Wang X, 2015 miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Research* 43(D1):D146-152.

Xie, B., Ding, Q., Han, H. and Wu, D., 2013. miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics*, p.btt014.

CHAPTER 3

3. Prognostic and predictive validation of microRNAs and their target genes as breast cancer biomarkers using an *in silico* approach

3.1 Introduction

A biomarker is defined as a clinical characteristic with the potential to distinguish normal biological processes from disease pathologies as well as to assess drug responses (Coccia, 2012). The ideal cancer biomarker should be easily accessible, sensitive enough to detect tumours at its different stages, and highly specific. Furthermore, the cancer biomarker must also have the ability for detection at an early stage or at disease recurrence as well as prediction of disease and treatment outcomes (Chatterjee & Zetter, 2005). In cancer, a large number of cellular, genetic, structural and metabolic components can serve as biomarkers as a wide variety of alterations occurs in the connecting networks thereby affecting overall cell growth and survival (Bahr *et al*., 2013). Hence, identification of biomarkers involves the study of intricate networks rather than the specific component of the pathway.

Based upon the utility such as cancer population screening, differential disease diagnosis, clinical staging and tumour growth measurement, assessment of disease and treatment outcomes, tumour or cancer biomarkers can be broadly divided as diagnostic, predictive and prognostic biomarkers (Kulasingam & Diamandis, 2008).

3.1.1 Diagnostic biomarkers

Diagnostic biomarkers are highly specific and sensitive biomarkers that are used to identify a given type of cancer both at an early stage and at different disease stages (Kulasingam & Diamandis, 2008). Despite current research advancements, a huge gap in the presence of such diagnostic biomarkers, specifically for early disease detection exists. While conventional diagnostic biomarkers largely involve histopathological variables, recent methodologies include the use of molecular biomarkers. These molecular biomarkers are detected using advanced proteomics analyses which are sensitive enough to detect small aberrations at transcriptional and post-transcriptional levels specific to the disease (Simpson *et al*., 2008). Recent studies have focused on the use of circulating microRNAs as diagnostic signatures in disease identification, including BC (Blenkiron *et al*., 2007; Iorio *et al*., 2005).

3.1.2 Prognostic biomarkers

Prognostic biomarkers are characterized as biomarkers with the ability to assess treatment and disease outcomes, specificity to detect cancer stage and disease remission prognosis (Panczyk, 2008; Paduch, 2003). Furthermore, the prognostic biomarkers concentration should directly implicate the disease stage and therapeutic outcomes of anti-cancer drugs for an individual (Bates, 1991). In a given tumour, prognostic biomarkers can be determined through gene expression polymorphism or mutation, alterations in DNA methylation or, recently through detection of circulating microRNAs and tumour cells (CTC) (Figure 3.1).

Mutations in DNA repair genes such as the BRCA family, ATM, p53, glutathione S-transferase family can be used as prognostic biomarkers as these genes enhance the likelihood for cancer susceptibility specifically BC (Cui *et al*., 2001; Synowiec *et al*., 2013). Similarly, G158A polymorphism has been reported to play a significant role in prostate cancer progression and hence can be used as an effective prognostic biomarker (Gong *et al.*, 2012; Samzadeh *et al*., 2012). Furthermore, the Genome-Wide Association Studies (GWAS) has enhanced the capability of quick detection of prognostic as well as pharmacogenetic markers by utilizing clinical information about genetic variations in disease conditions (Xu *et al*., 2010).

Extensive research regarding alterations in gene and miRNAs expression has further allowed the identification of prognostic biomarkers in aggressive disease conditions. In a study conducted by Liong *et al*., (2012). six genes, *CRTAM*, *CXCR3*, *FCRL3*, *KIAA1143*, *KLF12* and *TMEM204* were identified as prognostic biomarkers in prostate cancer. Furthermore, the expression levels of these genes can also determine the patient's segregation for disease aggressiveness. Similarly, several genes have been identified as prognostic biomarkers in BC for the assessment of treatment outcomes, cancer proliferation and invasiveness in BC using microarray, RT – PCR and *in silico* tools (van't Veer *et al*., 2002; Bueno-de-Mesquita, 2009; Nguyen *et al.*, 2012; Dabbs, 2011). Additionally, the presence of specific microRNAs has been correlated with several cancers (Qi *et al*., 2013; Jones *et al*., 2012; Wotschofsky *et al*., 2013).

DNA methylation has also been reported to act as a potential prognostic marker as hypermethylation patterns of gene encoding heat-shock protein 27 (*HSP-27*), *HSBP1*, can serve as a good prognostic biomarker (Vasiljevic *et al*., 2013). In-extension, circulating tumour cells (CTCs) that are detached from the tumour mass and migrates to the peripheral blood through the walls of blood vessels can also serve as a significant prognostic determinant (Hunter *et al*., 2008). In a study conducted by Wang *et al*., (2011) the presence of five or more CTCs was reported to be correlated with overall survival in prostate cancer patients.

3.1.3 Predictive biomarkers

Predictive biomarkers represent a treatment response prospective guiding clinicians to take therapeutic decisions (Walther *et al*., 2009; Voon *et al*., 2011). Commonly, somatic mutations in genes like *EGFR*, *KRAS*, *BRAF*, *PDGFRA*, *KIT*, *HER2*, *BCR-ABL*, and *EML4-ALK* has served as key predictive biomarkers. For instance, point mutations in the *KRAS* gene (codons 12, 13 and 61) and the *BRAF* gene (V600E) serve as standard predictive biomarkers for guiding treatment with targeted therapy against epidermal growth factor receptor (EGFR), using Cetuximab or Panitumumab in colorectal carcinogenesis (Walther *et al*., 2009; Lewandowska, 2012). Similarly, gene polymorphism in *CYP2D6*10/*10* (and *CYP2D6*5/*10*) and low concentrations of Endoxifen and 4-hydroxytamoxifen, the active metabolites of Tamoxifen has been correlated with weaker drug response in BC patients (Xu *et al*., 2008).

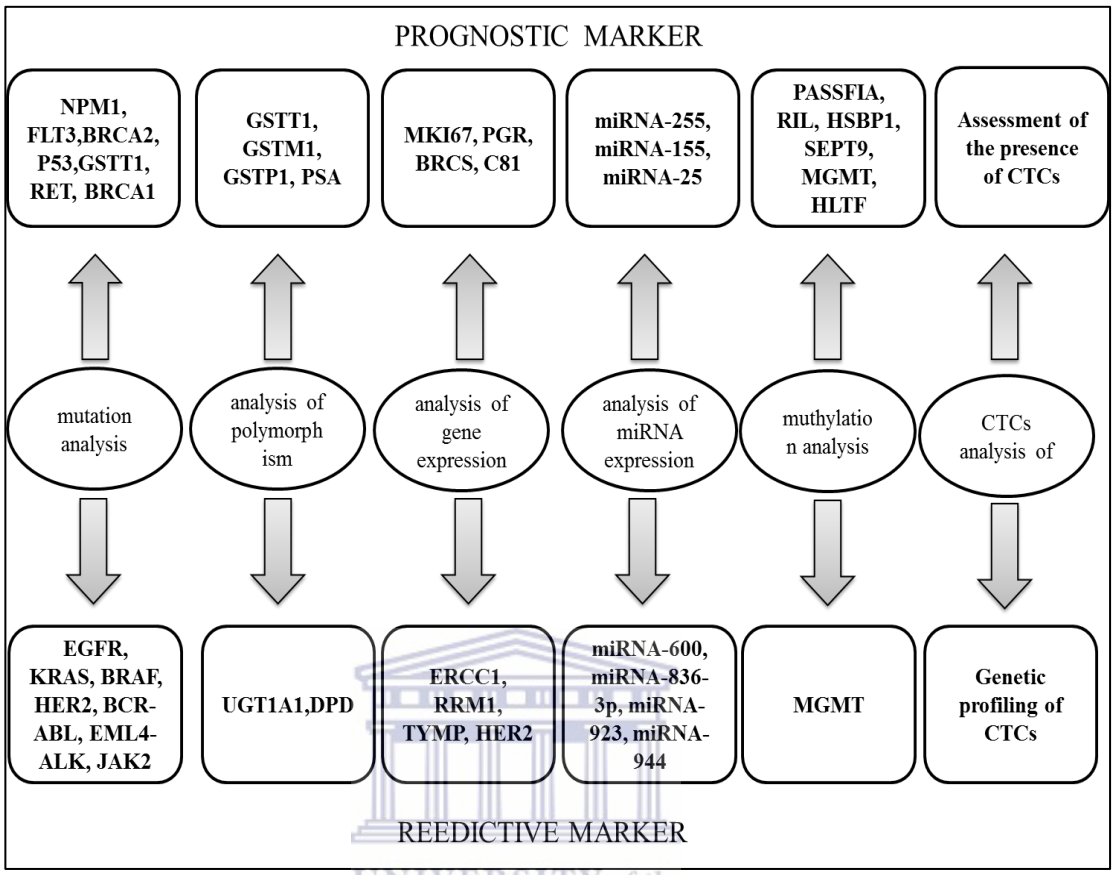


Figure 3.1: Prognostic and Predictive Biomarkers in Cancer Research (Adapted from Nalejska *et al*., 2014).

As depicted in Figure 3.1, both prognostic and predictive biomarkers share common features with varied specific gene targets. Similar to prognostic biomarkers expression analysis of DNA repair genes, microRNAs can also serve as predictive biomarkers in cancer. For example RNA, expression analysis of DNA repair gene *ERCC1*, as well as *mTOR* and *c-erb-B2* genes have been implicated as predictive biomarkers in hepatocellular carcinoma (Bassullu *et al*., 2012). However, due to the

absence of a large number of predictive biomarkers, studies on microRNAs as predictive biomarkers is gaining considerable importance. As previously discussed in Chapter 1 (section 1.10), a series of microRNAs involved in BC has been enumerated. In extension, determination of DNA methylation status and CTC count at different treatment time points also offers prediction of treatment outcomes (van den Bent *et al*., 2006; Hegi *et al*., 2004; Hegi *et al*., 2005; Karakousis *et al*., 2013).

3.2 Breast cancer biomarkers

Conventional BC biomarkers involve histological analysis of disease subtype and grade, lymph node metastases, and lymphovascular invasion. However, with the advent of high-throughput methods, several novel biomarkers have been reported with prognostic and predictive importance. To-date, only two biomarkers have been validated as per the American Society of Clinical Oncology's Tumour Marker Utility Grading System .These markers include the estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) (Hayes *et al.*, 1996; Harris *et al*., 2007). ER expression plays a crucial role as a predictive biomarker as the absence or presence of the ER provides information about endocrine treatment outcomes (Hammond *et al*., 2010). Furthermore, the expression of ER receptors strongly influences the expression of progesterone receptors (PgR). Several studies on adjuvant trials reported a strong prognostic value but mild predictive significance for PgR expression (Dowsett *et al.*, 2006). Oncogene HER2 was also identified as a prognostic biomarker in BC, as an up-regulated HER2 expression was found in BC

patients. Such patients are highly susceptible to disease recurrence and have a shorter overall survival (Mass *et al*., 2005). Ki67 is a nuclear non-histone protein that serves as a proliferation marker expressed during all phases of the cell cycle except the G0 phase and maximumally at M phase (Lopez *et al*., 1991). Ki67 has been well correlated with other BC biomarkers like cancer invasiveness, tumour grade, ER and HER-2 (Haerslev *et al.,* 1996; Nicholson *et al*., 1993; Rudolph *et al.,* 1999). Several research findings implicate Ki67 as a strong predictor of luminal A and B subtypes, chemotherapy response, disease recurrence and overall survival in BC patients (Cheang *et al.,* 2009; Faneyte *et al.,* 2003; Jones *et al*., 2009). Overexpression of cyclin D1 RNA has been implicated to be of prognostic importance in ER positive invasive BC (Ormandy *et al*., 2003; Bilalovic *et al*., 2005). Similarly, another cyclin protein i.e cyclin E has been reported to influence tumourigenesis and correlates with the progressive stage and grades of BC (Bortner & Rosenberg, 1997; Wingate *et al*., 2009). Furthermore, the role of cyclin E in the cell cycle was suggested to influence the response to chemotherapy and endocrine therapy in all cancers (Smith & Seo, 2000; Akli and Keyomarsi, 2004). In-addition, a large number of miRNAs has also been discovered as discussed in Chapter 1 (see section 1.13).

3.3 Cancer bioinformatics tools in prognostic and predictive validation of genes as biomarkers

Cancer bioinformatics involves the integration of systems biology, clinical science, omics-based technologies, bioinformatics and computational science to address

relevant challenges associated with early disease diagnosis, personalized therapies, and predictive prognosis of cancer patients. However, such methodological integrations for identification and validation of novel cancer biomarkers need computational tools and databases with high accuracy, specificity and applicability. Several bioinformatics tools has been developed to assess the predictive and prognostic utility of candidate biomarkers such as ITTACA, KMPlot, Recurrence Online, bc- GeneExMiner, GOBO, PrognoScan etc. (Aguirre-Gamboa *et al*., 2013). In the present study three tools (i.e Kaplan-Meier plot, SurvExpress and GOBO) was used and are described in detail below.

3.3.1 Kaplan-Meier plotter

The Kaplan-Meier plot is one of the popular statistical tools used for survival analysis after treatment. In this method, the fraction of living subjects post-treatment is estimated for a certain number of years (Goel *et al*., 2010). Other advantages of this analysis involve the assessment of sample loss from the study cohort in case of patient withdrawal and to compute the probable survival estimate by using multiple datasets which increase the statistical power significantly as reported by Györffy *et al*. (2010). Furthermore, the clinical association of the biomarker with patient survival can also be evaluated (Györffy *et al*., 2010). These survival plots are drawn through the Kaplan-Meier plotter, which is available online at http://kmplot.com/analysis/index.php?p=service&cancer=breast. An overall 22,277 genes from 10,188 cancer samples are assessed which include data from 4,142 breast,

1,648 ovarian, 2,437 lung, and 765 gastric cancer patients housed by this database. This huge database is maintained by a MySQL server and the retrieved data is calculated by using the R statistical tool (Győrffy *et al*., 2010). Each of the candidate genes can be plotted for its association with patient's survival and the output indicates the number of patients at risk, hazard ratio with 95 % confidence interval and logrank P (Győrffy *et al*., 2010).

3.3.2 SurvExpress

SurvExpress is a sophisticated gene expression database and bioinformatics tool equipped to generate survival analysis and risk assessment of cancer datasets. The tool can be accessed http://bioinformatica.mty.itesm.mx/SurvExpress and has the ability to validate survival and prognostic cancer biomarkers. The gene expression database comprises 20,000 samples from 20 different types of cancer along with 130 datasets. The candidate genes are used as input in the web interface and datasets to evaluate gene expression are selected from around 140 available datasets.

Aguirre-Gamboa *et al*. (2013) described SurvExpress as an online biomarker validation tool. They indicated the use of datasets to validate input genes and generate outcomes as Kaplan-Meier curves displaying risk groups, concordance index, and p-value of the log-rank testing equality of survival curves (Figure 3.2). In this figure, available clinical information related to risk groups, prognostic index, and outcome data gene expression values is shown as a heat map. The significance level (*p*-value)

using a t-test or f-test through box plots across risk groups and the relation between risk groups and prognostic indices is also shown in figure 3.2. A summary is given of the Cox fitting and the prognostic indices (Aguirre-Gamboa *et al*., 2013).

Overall, SurvExpress is one of the most reliable free online web tools to validate multi-gene biomarkers for gene expression in human cancers. The method provides quick outcomes at approximately one minute per dataset (Aguirre-Gamboa *et al*., 2013).

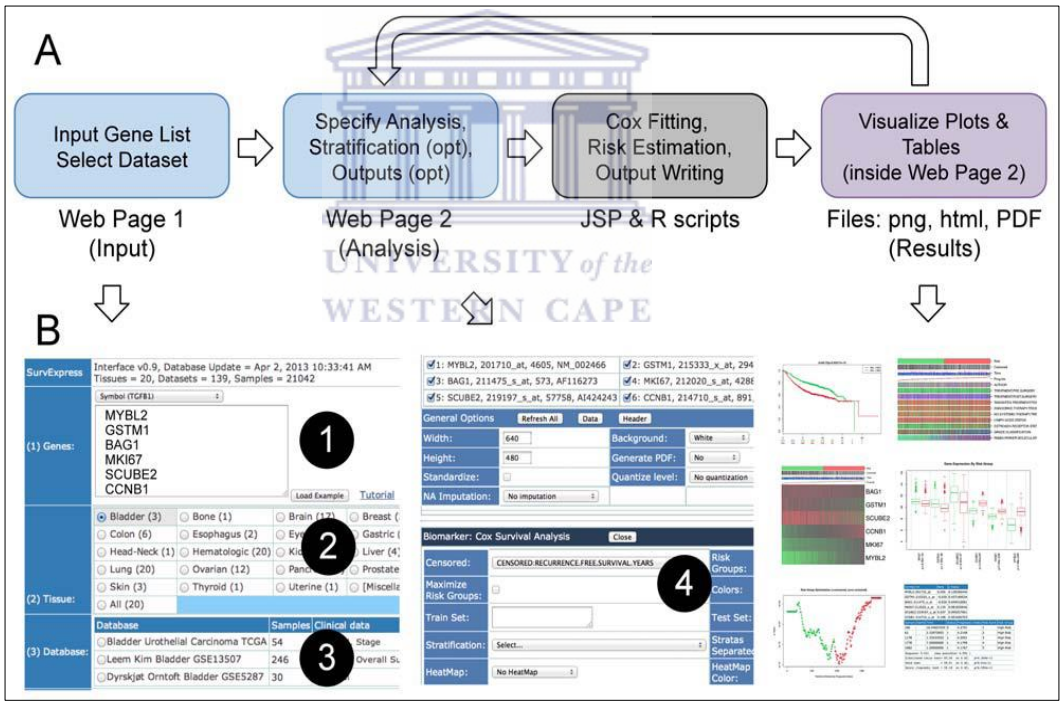


Figure 3.2: SurvExpress web tool (Adapted from Aguirre-Gamboa *et al*., 2013).

In figure 3.2, A is a schematic diagram of the SurvExpress workflow, while B shows snapshots of the interfaces tagging the required input fields. Number 1 is where the

list of genes are entered. Numbers 2 and 3 choose the dataset from more or less 140 available datasets. Number 4 shows the results from the input.

3.3.3 Gene expression-based outcome for breast cancer online (GOBO)

Gene Expression-Based Outcome for Breast Cancer Online (GOBO) is an online bioinformatics tool accessed at http://co.bmc.lu.se/gobo. GOBO helps in studying the differential expression of genes in the data repertoire generated through Affymetrix U133A microarrays for 1881 breast tumour dataset and 51 BC cell line datasets. These datasets can be interrogated for gene set analysis (GSA) in multiple breast tumours and cancer cell line subgroups, co-expressed genes (CG) assessment, and expression levels of single genes, sets of genes, or gene signatures in multiple subgroups i.e sample prediction (SP) in 1881-sample BC dataset (Ringnér *et al* ., 2011).

3.4 Aim and objectives

The main aim of this chapter was to determine the prognostic and predictive values of five novel microRNAs and their associated (*ACTN4*, *TMED10*, *STY9*, *GOPC*, *VEGFA*, *CD2AP*, *ANP32F*, *HIP1* and *AP1G1)* as putative biomarkers in BC using different bioinformatics tools. The study was conducted using the following objectives-

i) To validate the candidate microRNAs for their prognostic/predictive impact using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2.

ii) To validate the candidate genes regulated by these microRNAs for their prognostic/predictive impact using two datasets from SurvExpress.

iii) To analyze the predictive importance of individual biomarker candidate genes using the Kaplan-Meier plotter.

iv) To determine the prognostic/predictive significance of the candidate genes in different BC types using GOBO.

3.5 Materials and methods

3.5.1 Validating the candidate miRNAs for their prognostic/predictive impact using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2

Initially, the five identified microRNAs were analysed for their predictive and prognostic impact using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2.

3.5.1.1 SurvMicroRNA

The SurvMicroRNA tool can be accessed at http://bioinformatica.mty.itesm.mx:8080/Biomatec/ Survmicro.jsp to assess microRNA-based prognostic signatures for BC outcomes by multivariate survival

analysis. All microRNAs (five) unique identifiers were used as input in the space provided for microRNA list and the "breast" option was selected as tissue type. In the option for "duplicated genes", the default setting was selected. The software assessed the five microRNAs using two different broad datasets by performing a quantile normalization compiled by SurvExpress i.e. DATA 1- Breast Invasive Carcinoma (Illumina GA) the Cancer Genome Atlas (TCGA) dataset using 322 samples and DATA 2-Buffa Camps Breast GSE22216 Dataset with 210 samples.

3.5.1.2 miRpower

Similarly using miRpower, the five microRNAs provided no output http://kmplot.com/analysis/index.php?p=service&start=1. miRpower is a web-tool to validate survival-associated microRNAs utilizing expression data from 2,178 BC patients.

3.5.1.3 Kaplan-Meier plotter

Furthermore, analysis was performed using the Kaplan-Meier plotter, which is accessed at http://kmplot.com/analysis/index.php?p=service&cancer=breast. All the candidate microRNAs were used as an input and Kaplan-Meier plots were drawn.

3.5.1.4 PROGmiRV2

The five candidate microRNAs were analyzed using PROGmiRV2 found at http://xvm145.jefferson.edu/progmir/index.php. For cancer type 'breast invasive carcinoma' (BRCA) were selected. All microRNAs were used as input and survival measure "metastasis" was selected for outcomes.

3.5.2 Validating the candidate genes regulated by these miRNAs for their prognostic/predictive impact using two datasets from SurvExpress

The nine identified candidate genes (*ACTN4*, *TMED10*, *STY9*, *GOPC*, *VEGFA*, *CD2AP*, *ANP32F*, *HIP1* and *AP1G1)* regulated by the 5 microRNAs were assessed for their prognostic and predictive ability using the SurvExpress tool. The SurvExpress tool was launched at http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp and the list of nine genes were used as input using the gene list option. For the tissue type option "breast" was selected whereas default setting was used for the duplicated genes option. The candidate genes (nine) were analyzed using two broad datasets; (1) One comprising nine datasets from the same platform / methods that were compiled by seven authors comprising 1,574 samples (DATA 1) and (2) meta-bases generated through a quantile normalization compiled by SurvExpress comprising 1901 samples (DATA 2). DATA 1 and DATA 2 are recurrent clinical data using meta-analysis (Aguirre-Gamboa *et al*., 2013). The input genes were analyzed using the output "Recurrence Free Survival

Years" and the output was saved as a PDF file. Outputs are given as Kaplan-Meier plots, heat maps and box plots.

3.5.3 Analyzing the predictive importance of individual biomarker candidate genes using the Kaplan-Meier plotter analysis/database

The analysis was performed using the Kaplan-Meier plotter. All microRNAs regulated genes were individually assessed for their involvement in BC. For this web tool, each gene was used as an input for the option "gene space" whilst "breast" was utilized in the option 'cancer type'. For all remaining sections the default settings were used to draw Kaplan-Meier plots. The resultant output was saved as a PDF file.

3.5.4 Determining prognostic/predictive significance of the candidate genes in different breast cancer types using GOBO

The prognostic importance of the genes was evaluatedusing the GOBO database which was accessed at http://co.bmc.lu.se/gobo. The genes were analyzed using the GSA-Tumours application in GOBO where gene symbols were used as input and the option of different tumour types were selected. In the default settings full year censoring was selected.

3.6 Results

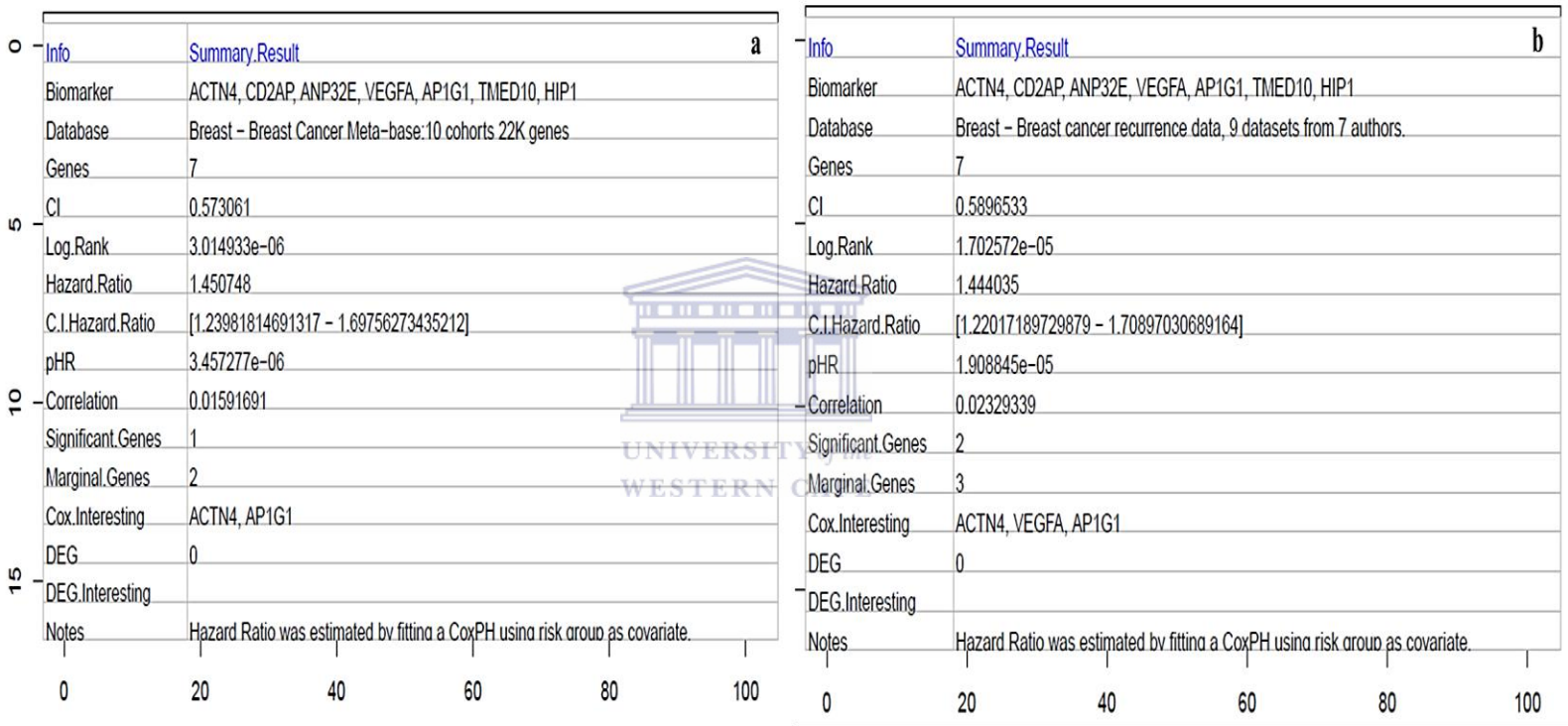
3.6.1 Analysis of microRNAs by using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2 databases

All five microRNA symbols when placed individually as an input into the selected tools, displayed no predictive or prognostic output for BC.

3.6.2 Analysis of microRNA candidate genes by using SurvExpress database

Table 3.1 demonstrates the output of the SurvExpress tool where nine breast cancer candidate genes regulated by five microRNAs were analyzed for their prognostic value. The resulting two datasets: DATA 1 and 2 only reflect on 7 of the 9 genes queried because there was no information on two of the genes in these datasets. Three (*ACTN4*, *VEGFA* and *AP1GI*)of seven genes were identified as significant prognostic markers based on their p-values ($p < 0.05$) by DATA 1 which contains 9 datasets with an overall sample size of 1,857 as compiled by different authors. However, DATA 2 identified two genes (*ACTN4* and *AP1GI*) as significant prognostic importance comprising 1,901 samples constructed as meta-analysis. Both datasets identified similar genes except *VEGFA*.

Table 3.1: Significant prognostic genes identified from two datasets; (a) DATA 2 and (b) DATA 1 using SurvExpress



a

Info	Summary.Result
Biomarker	ACTN4, CD2AP, ANP32E, VEGFA, AP1G1, TMED10, HIP1
Database	Breast – Breast Cancer Meta-base:10 cohorts 22K genes
Genes	7
CI	0.573061
Log.Rank	3.014933e-06
Hazard.Ratio	1.450748
C.I.Hazard.Ratio	[1.23981814691317 – 1.69756273435212]
pHR	3.457277e-06
Correlation	0.01591691
Significant.Genes	1
Marginal.Genes	2
Cox.Interesting	ACTN4, AP1G1
DEG	0
DEG.Interesting	
Notes	Hazard Ratio was estimated by fitting a CoxPH using risk group as covariate.

b

Info	Summary.Result
Biomarker	ACTN4, CD2AP, ANP32E, VEGFA, AP1G1, TMED10, HIP1
Database	Breast – Breast cancer recurrence data, 9 datasets from 7 authors.
Genes	7
CI	0.5896533
Log.Rank	1.702572e-05
Hazard.Ratio	1.444035
C.I.Hazard.Ratio	[1.22017189729879 – 1.70897030689164]
pHR	1.908845e-05
Correlation	0.02329339
Significant.Genes	2
Marginal.Genes	3
Cox.Interesting	ACTN4, VEGFA, AP1G1
DEG	0
DEG.Interesting	
Notes	Hazard Ratio was estimated by fitting a CoxPH using risk group as covariate.

Kaplan-Meier plots indicating the log-rank test of differences, the hazard-ratio, and the concordance indices between the risk groups were also constructed from both datasets as illustrated in Figure 3.3. All seven genes showed significance ability in the prediction of prognostic outcomes of BC patients. The subjects with a higher risk (red) will show high probability of relapse in comparison to patients with lower risk (green) from both databases.

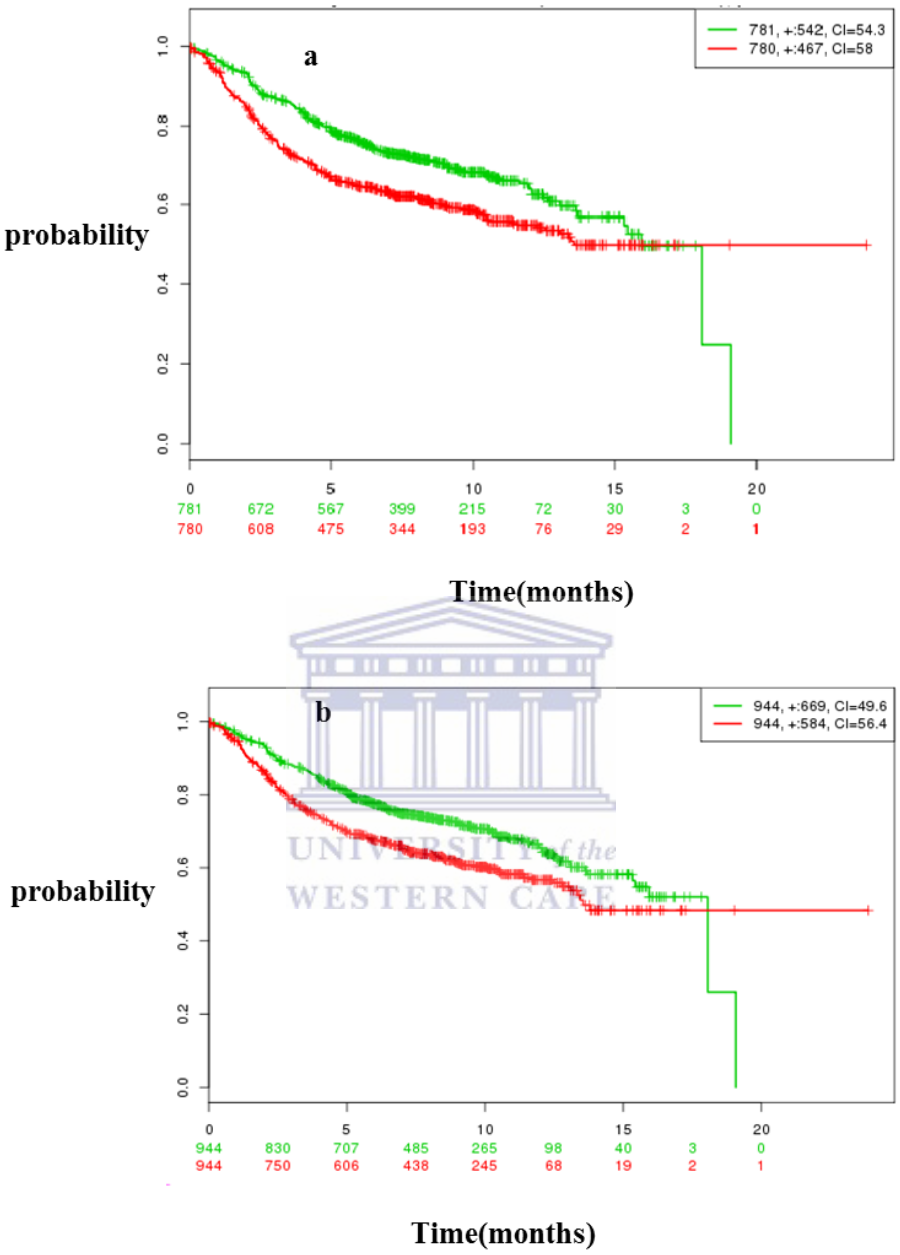


Figure 3.3: Kaplan-Meier analysis for all microRNA target genes for breast cancer. By class: treatment/ pre-surgery (overall betas) concordance index = 58.97, Log−Rank Equal Curves p=1.703e−05, R2=0.023/0.993 Risk Groups Hazard Ratio = 1.44 (conf. int. 1.22 ~ 1.71), p=1.909e−05 a) DATA 1. b) DATA 2.

Figure 3.4 represent the heat map generated from the two datasets based on the prognostic ability of each of the seven target genes. The colour codes in the heat map demonstrates the expression level of each target genes against samples in the two datasets based on their prognostic index. From the DATA 1, four genes (*CD2AP*, *TMED10*, and *ACTN4*) were highly expressed in the high risk group and the *HIP1* gene was highly expressed in the low risk group. The remaining genes were found to be differentially expressed in both the groups. Similarly, from DATA 2 two genes, (*TMED10* and *ACTN4)* were highly expressed in the high risk group and *HIP1* gene was highly expressed in low risk group. The remaining genes were differentially expressed in both high and low risk the groups.

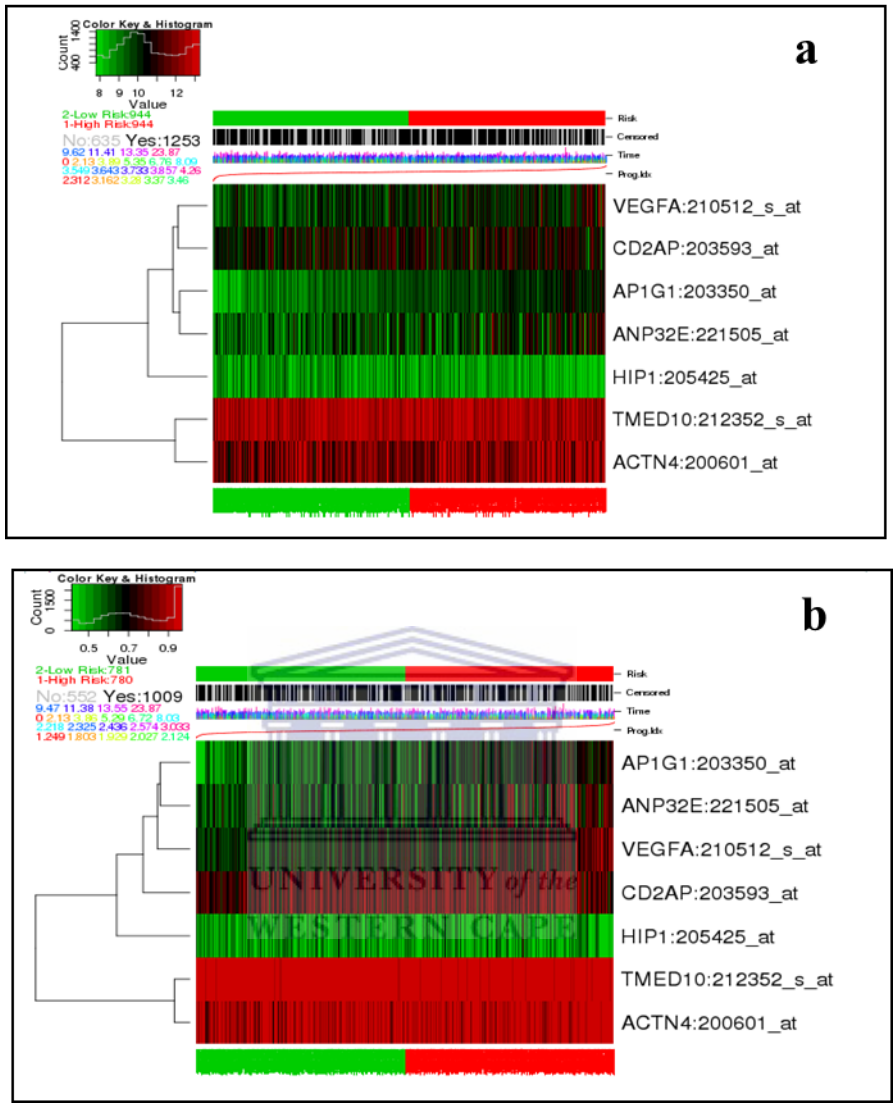


Figure 3.4: Expression profile and gene ranking of targeted genes based on the prognostic index. (a) Expression profile from the datasets compiled by SurvExpress comprising 1,857 samples(DATA 1). (b) Expression profile from dataset of nine databases compiled by seven authors comprising 1,901 samples(DATA 2)

Figure 3.5 demonstrate the box plot graph output from SurvExpress analysis indicating the expression level of the genes plotted along risk groups obtained in the analysis. From the box plot, it was assessed whether gene expression levels differed between the high risk (red) and low risk groups (green) as well as the level of significance of expression (p-value < 0.05) using a t-test. From DATA 1, it was derived that all genes are differentially expressed in both the high and low risk samples, except *TMED10* which was found to be uniformly expressed in both the groups. Furthermore, the prognostic significance of *ANP32E*, *VEGFA* and *AP1GI* largely remains the same between both the groups where all these genes show similar expression levels. Based on DATA 2, all the genes were found to be differentially expressed in both the groups. Overall from DATA 1 seven genes were differentially expressed whilst five genes from DATA 2 (p-value < 0.05).

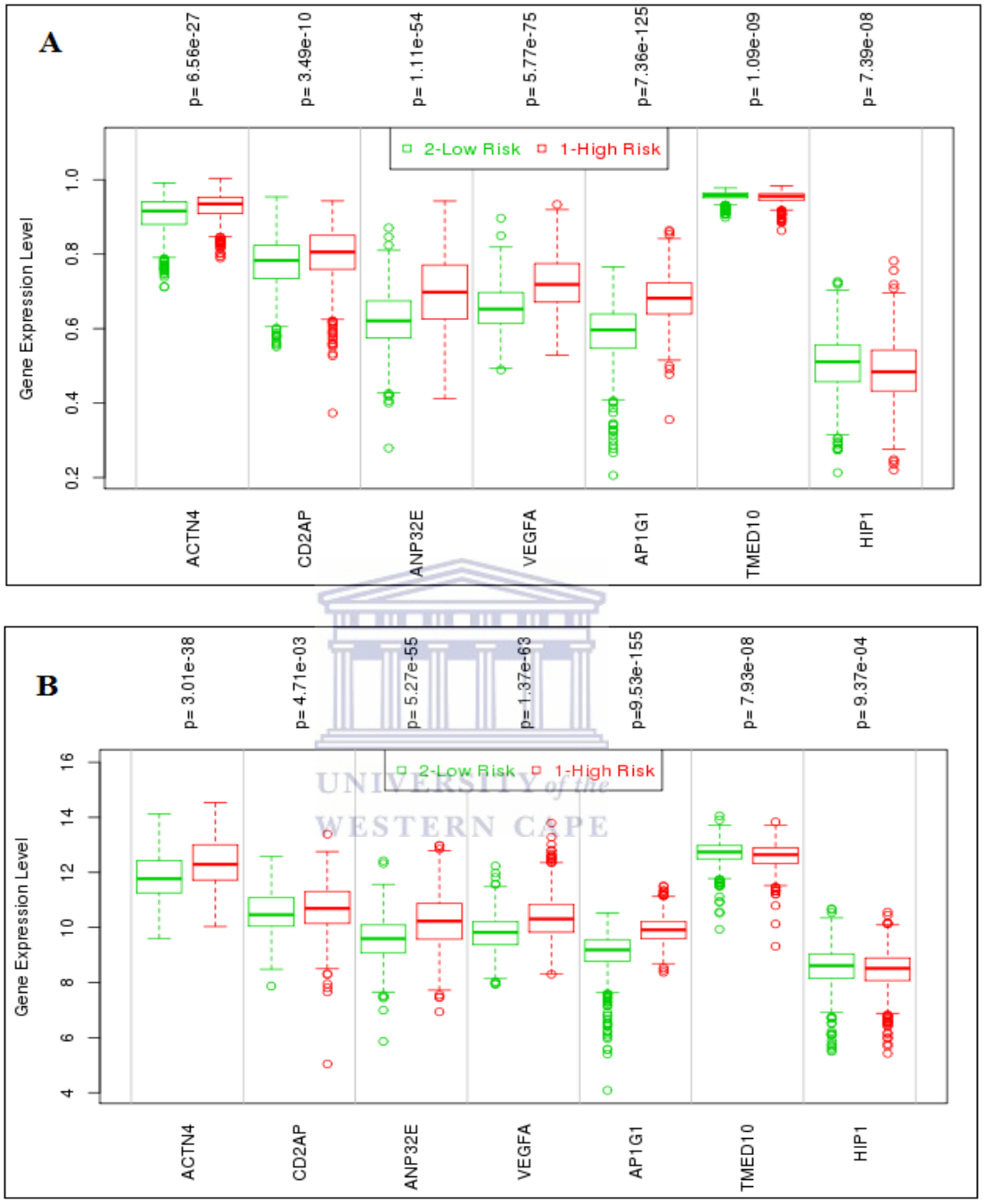


Figure 3.5: Gene expression analysis of microRNA target genes using box plots in low and high risk groups from (a) DATA 1 and (b) DATA2.

3.6.3 Analysis of microRNA target genes using Kaplan-Meier plotter database

The prognostic importance of each of the candidate biomarker genes for BC outcome was studied using the Kaplan-Meier plotter and the results of individual genes are displayed in Figure 3.8,. These figures showed differential survival of risk groups based on the gene expression levels of the target genes. While 7 of the 9 target genes (*CD2AP*, *VEGFA*, *HIP1*, *SYT9*, *GOPC*, *AP1G1*, and *ANP32)* showed significant correlation with BC outcomes (p-value < 0.05), the remaining two gene (*ACTN4* and *TMED10)* was found to be of poor prognostic importance (p-value >0.05). Moreover, while *AP1G1* and *VEGFA* matched with the results of SuvExpress, *ACTN4* was not found to be of any prognostic relevance by this tool. Furthermore, the remaining genes (*CD2AP*, *HIP1*, *SYT9*, *GOPC*, and *ANP32*) were also reported to be of prognostic importance using the Kaplan-Meier plotter alone and not in the SurvExpress analysis.

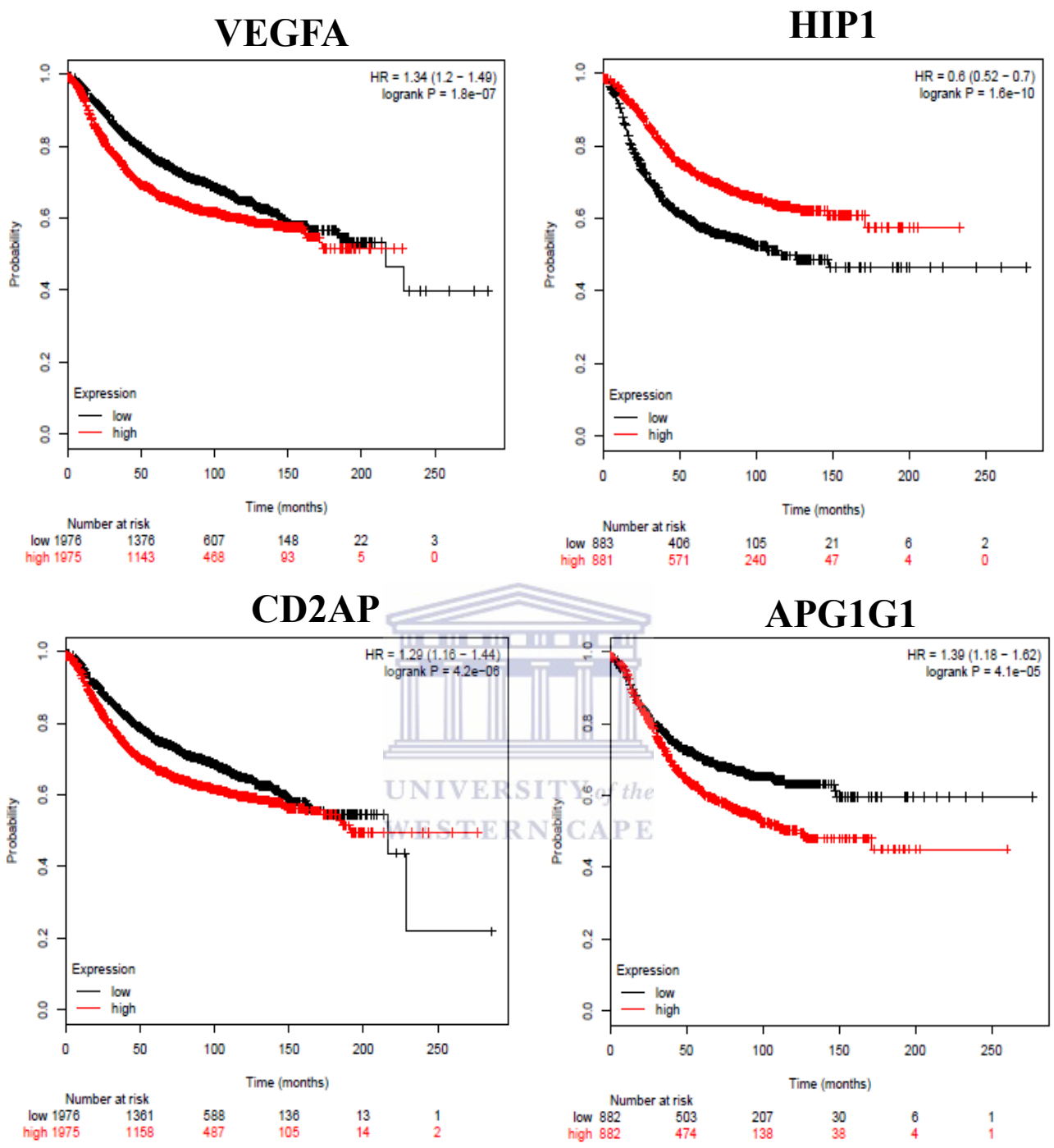


Figure 3.6: Survival risk curves for several target genes (*VEGFA, HIP1, CD2A* , and *AP1G1)* Low and high risks are drawn in black and red respectively. The *p*-value was shown in right hand corner of each curve ($p \leq 0.05$ is significant).

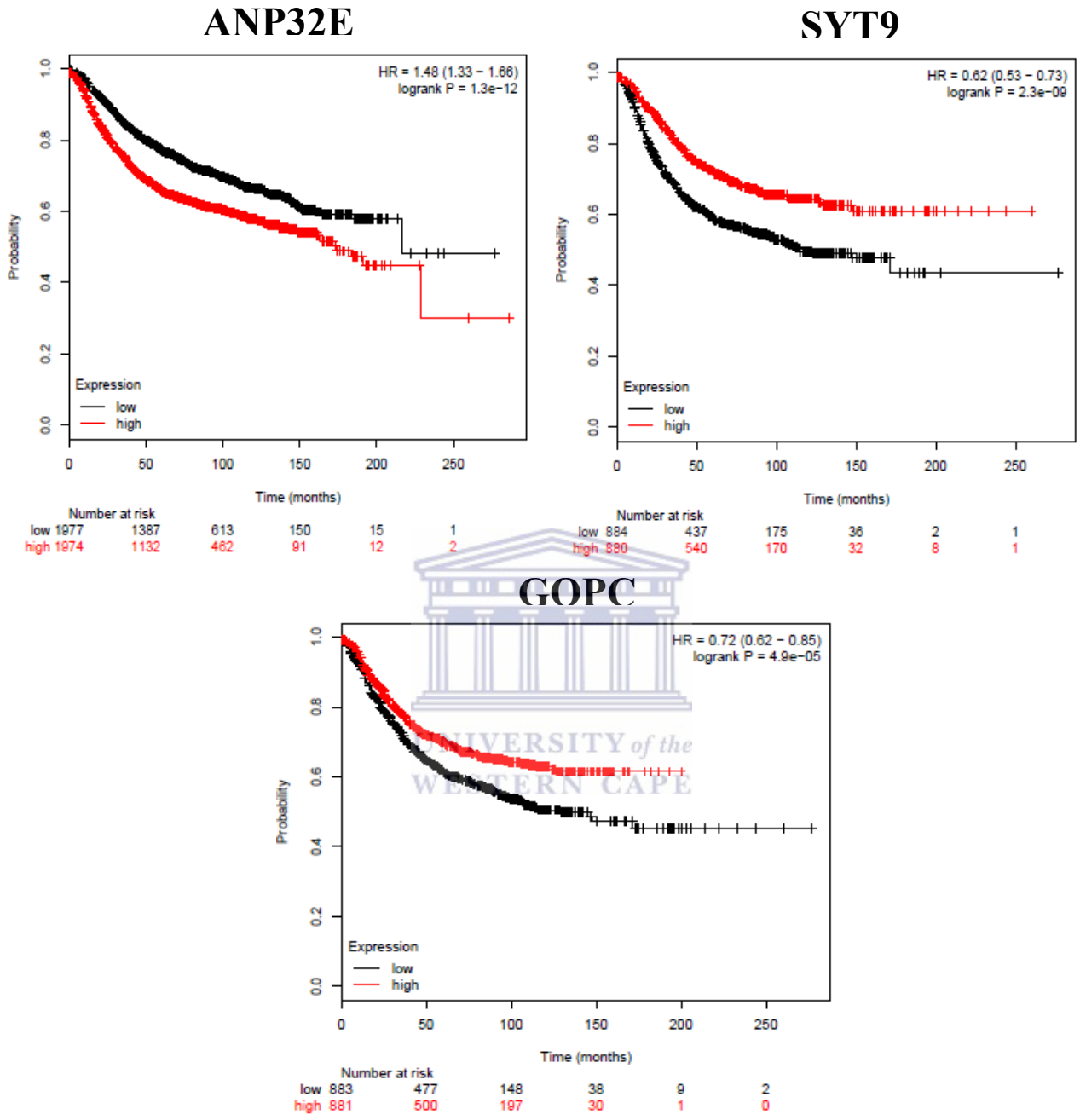


Figure 3.7: Survival risk curves for target genes (*ANP32E*, *SYT9* and *GOPC).* Low and high risks are drawn in black and red respectively. The p-value was shown in right hand corner of each curve ($p \leq 0.05$ is significant).

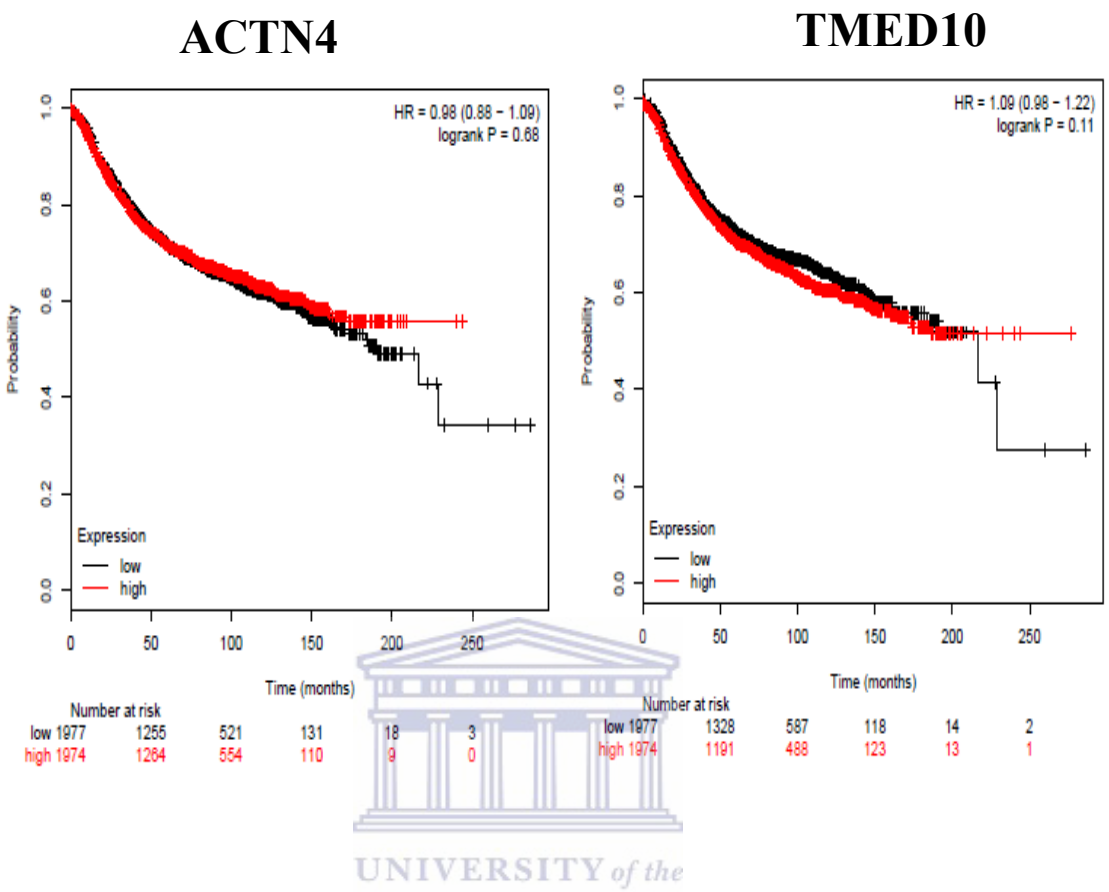


Figure 3.8: Survival risk curves for target genes (*ACTN4* and *TMED10)*. Low and high risks are drawn in black and red respectively.The *p*-value was shown in right hand corner of each curve ($p \geq 0.05$ is significant).

3.6.4 Analysis of microRNA target genes by gene expression-based outcome for breast cancer online (GOBO)

The prognostic ability of candidate biomarker genes and their impact on survival outcomes in low-, intermediate- and high-risk patients in different BC subtypes – estrogen receptor positive all tumour subtypes or luminal B tumour subtype; estrogen receptor negative all tumour subtypes or luminal B tumour subtype or ERBB2 tumour

subtype were evaluated through GOBO and the results are displayed in Figure 3.8. The survival curve was plotted as percentage distant metastasis-free survival (DMFS) against time in years. The survival curve expressed a significant prognostic importance of all the nine candidate genes in predicting the outcomes in all estrogen receptor positive BC subtypes and estrogen receptor negative luminal B tumour subtype (p–value < 0.05). On the contrary, the outcomes of remaining BC subtypes cannot be predicted using these candidate biomarker genes as the p-value was found to be highly insignificant (p–value > 0.05).

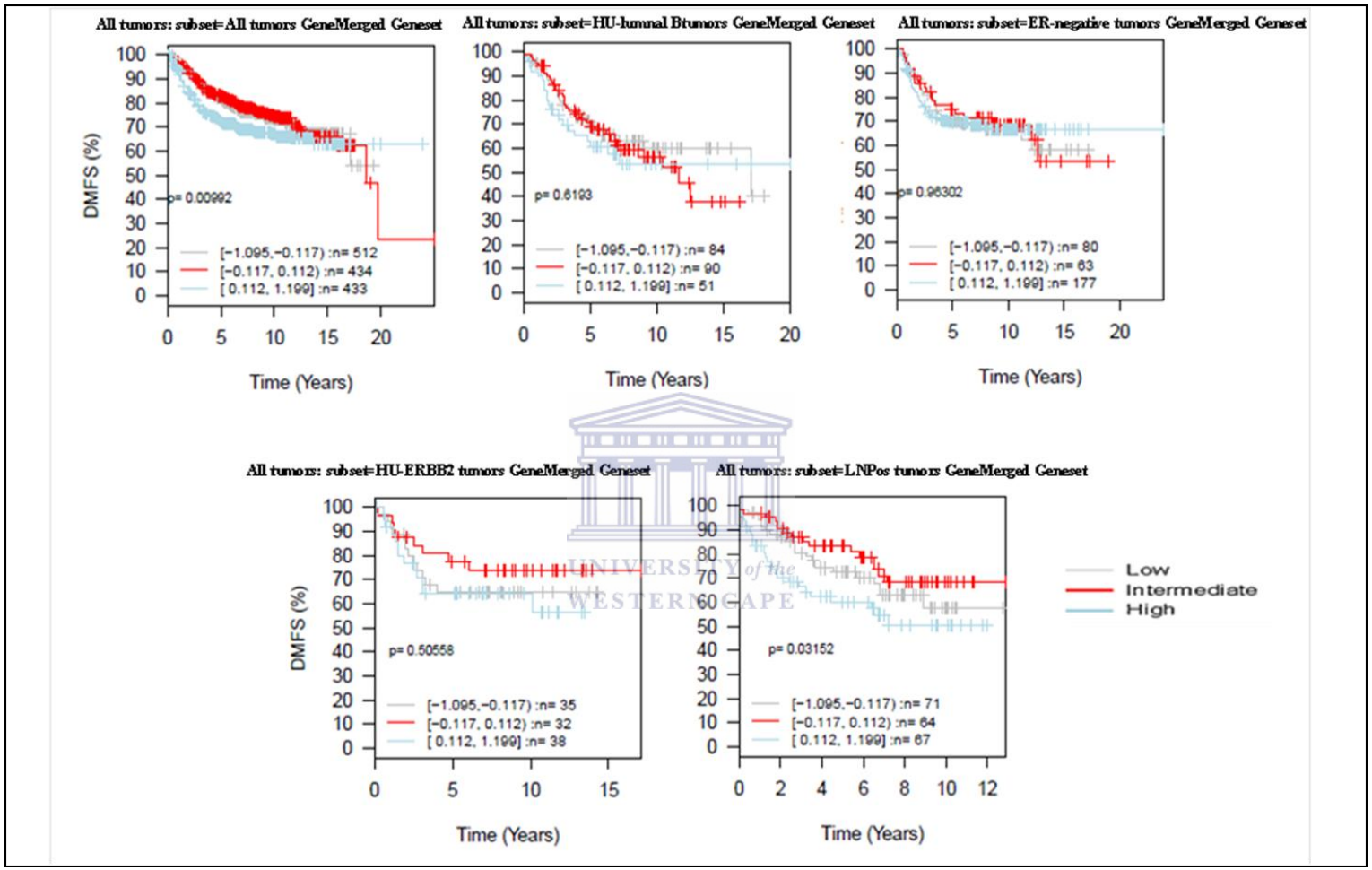


Figure 3.8: Survival curve for different breast cancer subtypes (a) ER all tumour subtype (b) ER luminal B tumour subtype (c) ER all tumour subtype negative tumour (d) ER- ERBB2 tumour subtype (e) ER- luminal positive tumour.

3.7 Discussion

Breast cancer involves dysregulation of multiple networks encoded by a series of microRNAs and proteins. Identification of gene regulation by microRNAs has considerable predictive and prognostic significance for disease outcomes. With the advent of high-throughput methods, a large number of datasets containing the gene expression profile of multiple cancer subtypes has been developed. Such databases can be explored by *in-silico* tools to validate the candidate genes as prognostic or predictive biomarkers in different BC subtypes. The selected genes can further be validated in clinical cohorts, which therefore help in identifying the specific genes as biomarkers from a huge reservoir in a fast and cost effective manner. In the previous chapters, five microRNAs regulating nine genes were identified as candidate biomarkers from a huge datasets and their prognostic and predictive significance was studied in this chapter.

3.7.1 Prognostic/predictive value of the microRNAs using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2

Initially each of the individual microRNAs were analyzed for their predictive and prognostic importance using SurvMicroRNA, miRpower, Kaplan-Meier plotter, and PROGmiRV2. However, none of the web tools produced results for these five microRNAs suggesting their novelty in BC. Hence, their predictive and prognostic impact was assessed indirectly by analyzing the prognostic outcomes of the nine genes regulated by these microRNAs.Due to the novelty of the five microRNAs in

BC as per the criteria used for microRNA identification, the absence of results for these microRNAs in terms of their prognostic value lends more evidence for their potential novelty in BC.

3.7.2 Prognostic/ predictive -value of the microRNA target genes by using SurvExpress, Kaplan-Meier plotter and GOBO

As indicated in Chapter 2, the first candidate microRNAs are symbolized as MiR1 was found to regulate gene *HIP1*. In the initial SurvExpress tool output the prognostic importance of this gene was not shown utilizing datasets DATA 1 and DATA 2. However, as represented by the heat map and box plot output of SurvExpress, the *HIP1* gene was found to be highly expressed in low-risk BC patients. From the Kaplan-Meier plotter results, the *HIP1* gene was found to significantly correlate with prognosis and differential survival of BC patients. This suggested that the target gene and its regulating microRNA, MiR1 has considerable prognostic importance. It can be deduced that low expression of this gene can affect the prognostic outcome in high-risk patients.

MiR2 regulated the expression of gene *SYT9* whil no prognostic significance of this gene was obtained in the SurvExpress analysis. The Kaplan-Meier plotter outcomes revealed a fair association of this MiR2 in BC prognosis and survival outcomes thereby suggesting the importance of MiR2.

MiR3, another signature microRNA identified in the previous chapter have been shown to regulate the expression of two genes – *TMED10* and *AP1G1*. Using SurvExpress analysis, *AP1G1* was found to be commonly identified as a significant prognostic biomarker in the initial analysis of both datasets, *TMED10* was found to be highly expressed in the high risk group from the heat map and box plot output using the two datasets. Furthermore, *AP1G1* prognostic importance was reproduced in the Kaplan-Meier plotter analysis where the candidate biomarker showed good prediction of differential survival in BC patients. In a previous published report, *AP1G1* has been identified as one of the prognostic signature genes among eleven genes in tumour samples collected from triple negative breast cancer (TNBC) and lymph node negative BC subtypes patients (Liu *et al*., 2014). Furthermore, these patients were not subjected to adjuvant chemotherapy and hence the biomarker gene can be useful in predicting the drug response in such patients (Liu *et al*., 2014). Although, the prognostic importance of the gene has been studied, outcomes related to its regulating microRNA remains unnoticed.

Similarly, MiR4 regulates the gene expression of *VEGFA*, *ANP32E* and *CD2AP*. While SurvExpress analysis identified *VEGFA* as a significant prognostic marker and high expression of *CD2AP* in the high risk group, Kaplan-Meier plotter outcomes demarcated all three genes for their prognostic importance and prediction of differential survival among BC patients. *VEGFA* was reported to be aggressively expressed in untreated metastatic BC and the lymph node negative subtype (Liu *et al*., 2014). Another microarray study demonstrated correlation between increased

VEGFA expression and poor survival (Ghosh *et al*., 2008). Furthermore, another study revealed the association of *VEGFA* with adjuvant therapy resistance in postmenopausal BC patients (Ryden *et al*., 2005). Hence, identification of *VEGFA* in BC subtypes can help in predicting the disease aggressiveness as well transition from benign to malignant types.

Finally, MiR5 was found to regulate two genes (*ACTN4* and *GOPC)* as described in previous bioinformatics analysis (see section 3.6.2). All the SurvExpress tool outputs identified *ACTN4* as a significant prognostic marker in high risk group. On the contrary, Kaplan-Meier plotter outcomes identified the prognostic importance of GOPC alone in prediction of differential survival of BC patients. *ACTN4* acts as a regulatory gene in signal transduction, nuclear translocation, and gene expression. In BC, *ACTN4* may serve as a potential biomarker as it is reported to act as a transcription co-activator for cell proliferation and adjuvant endocrine therapy resistance (Hsu & Kao, 2013).

Using GOBO analysis, all nine candidate genes showed significant prognostic importance in predicting the survival outcomes in low, intermediate and high risk patients specifically with estrogen receptor positive, all tumour subtypes, and estrogen receptor negative luminal B tumour subtype. In a recent study where survival analysis was performed based on the ER/PR/HER2 subtypes and tumour grade in 123,780 cases of invasive BC from stages 1–3, registered under California Cancer Registry, a better survival and adjusted mortality was reported in all the ER+

subtypes in comparison to ER− subtypes, except at disease stage 3. In Stage 3, all ER+ subtypes showed poor survival and the study suggest that ER may serve as a crucial factor in survival outcomes (Carol & Caggiano, 2014). Hence, the significant correlation between the nine selected genes and better survival outcomes in estrogen receptor positive all tumour subtypes obtained in the present study may have promising value as biomarkers, at least in these subtypes. These results will help in identification of a group of ER+ BC patients who are resistant to endocrine treatment. In addition to ER positive subtypes, the present study results also displayed significant association between better prognostic outcomes of selected biomarker genes in estrogen receptor negative luminal B tumour subtype. This BC subtype displays a highly poor prognosis as well as severe disease characteristics such as reduced hormone receptors, up-regulated proliferation markers and higher histological grade of the disease (Ades *et al*., 2014). Hence, identification and validation of biomarker genes with improved survival outcomes and better treatment prognosis can greatly assist treatment and response for luminal B subtype BC patients. Similar studies was performed by Matboli *et al*., (2014) where differentially higher expression of Histidine-rich glycoprotein RNA (HRG) was identified as one of the prognostic biomarkers in basal and grade 2 BC subtypes using GOBO and reverse transcriptase (RT-PCR) (Matboli *et al*., 2014).

Overall, the results presented here indirectly suggests the importance of the candidate microRNAs as prognostic and predictive signatures in BC outcomes as well as their regulated genes. In a recent report, Bertoli *et al*. (2015) has summarized several BC

microRNAs that has been identified as diagnostic, prognostic and predictive signatures Several studies attempted to demarcate the diagnostic signature microRNAs that can distinctly differentiate BC from normal tissues. For instance, whilst 13-miRNAs was identified for diagnostic importance by Iorio *et al*.(2005), 133 signature microRNAs were reported by Blenkiron colleagues in breast tumour tissues (Blenkiron *et al*., 2007). In a study conducted by Blenkiron *et al*.(2007*)* , of the 309 microRNAs for a total of 93 BC subtypes, it was shown that 9 miRNAs can effectively demarcate luminal A from luminal B subtype . Other studies conducted by Lowery et al.(2009) and Volinia et al.(2006) identified 15-microRNAs and 17-microRNAs as predictive signatures in ER, PR and HER2 positive BC subtypes respectively Similarly, multiple reports identifying several circulating and non-circulating microRNAs with diagnostic or prognostic implications have been proposed. Additionally, microRNAs that can predict the therapeutic outcomes against the common chemotherapeutics like antracyclines, selective ER modulators, 5-fluorouracil (5-FU), and cyclophosphamide has also been reported (Bertoli *et al*., 2015). The present study outcomes has potentially added new microRNAs with prognostic importance to the BC list specifically for estrogen receptor positive all tumour subtypes and estrogen receptor negative luminal B tumour subtype.

Despite the achievement of candidate microRNAs as prognostic and predictive signatures in BC subtypes, a large amount of variability was obtained in the output of different bioinformatics tools used in the current study. The variability in outputs of two datasets of SurvExpress tool was perhaps due to the use of different approaches

in data compilation and the same can be corroborated by a previous study where variations were obtained in biomarkers performance using different database (Aguirre-Gamboa *et al.,* 2013). Furthermore, the differences in output from two datasets may also be due to the variability in cancer types and subtypes, patient cohorts and their tumour characteristics, microarray and mathematical methods of analysis (Fulford *et al.,* 2007; Fan *et al.,* 2006). Whilst Kaplan-Meier plotter uses a database of BC patients undergoing multiple treatments, SurvExpress uses the data from generated qRT – PCR analysis from a few BC cell lines on the other hand. Therefore, databases generated from clinical samples and cell lines will contain obvious disparities as a huge gap exists between the disease representation by clinical samples and cell lines (Holliday &Speirs, 2011). Nevertheless, the methodological approach for validation of prognostic biomarkers were used previously and significant correlations between expression of putative gene and BC survival was noticed in specific cancer subtypes (Romagnoli *et al*., 2012). Therefore, it is believed that the present methodology will also succeed in obtaining novel biomarker microRNAs and their genes which can effectively predict the disease and treatment outcomes with large accuracy, sensitivity and specificity in multiple BC subtypes.

To conclude, results present in this study represented the prognostic and predictive importance of five microRNAs and their regulated genes using SurvExpress, Kaplan-Meier and GOBO analysis, which could be promising when used to predict disease, treatment and survival outcomes in different BC subtypes. Although variance in gene selection for its prognostic value was obtained for each of the tools, due to the

variable nature and source of information available in different datasets, the study owes its importance to the fact that cancer is a heterogeneous disease involving intricate networks of genes. Hence, a multitude of genes should be studied to extract maximum benefits. Since all the regulated genes were more or less found to be differentially expressed in BC, they all have the potential to serve as prognostic or predictive signature in BC. Furthermore, the clear relationship between the expressions of the candidate genes in BC subtypes further enhances the study utility in accurate prediction of survival, disease and treatment outcomes in populations like estrogen receptor positive all tumour subtypes and estrogen receptor negative luminal B tumour subtype. These genes will be further validated for their differential expression using bioinformatics tools (Chapter 4).

3.8 References

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), pp.2794-2803.

Aguirre-Gamboa, R., Gomez-Rueda, H., Martínez-Ledesma, E., Martínez-Torteya, A., Chacolla-Huaringa, R., Rodriguez-Barrientos, A., Tamez-Pena, J.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), p.e74250.

Akli, S. and Keyomarsi, K., 2004. Low-molecular-weight cyclin E: the missing link between biology and clinical outcome. *Breast Cancer Research*, 6(5), p.188.

Barh, D., Zambare, V. and Azevedo, V. eds., 2013. *Omics: applications in biomedical, agricultural, and environmental sciences.* CRC Press, p 713.

Bassullu, N., Turkmen, I., Dayangac, M., Korkmaz, P.Y., Yasar, R., Akyildiz, M., Yaprak, O., Tokat, Y., Yuzer, Y. and Dogusoy, G.B., 2012. The predictive and prognostic significance of c-erb-B2, EGFR, PTEN, mTOR, PI3K, p27, and ERCC1 expression in hepatocellular carcinoma. *Hepatitis monthly*, 12(10 HCC).

Bates, S.E., 1991. Clinical applications of serum tumor markers. *Annals of Internal Medicine*, 115(8), pp.623-638.

Bertoli, G., Cava, C. and Castiglioni, I., 2015. MicroRNAs: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. *Theranostics*, 5(10), p.1122.

Bilalović, N., Vranić, S., Bašić, H., Tatarević, A. and Selak, I., 2005. Immunohistochemical evaluation of cyclin D1 in breast cancer. *Croatian Medical Journal*, 46(3), p p (382-388).

Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.F., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O. and Tavaré, S., 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome biology*, 8(10), p.R214.

Bortner, D.M. and Rosenberg, M.P., 1997. Induction of mammary gland hyperplasia and carcinomas in transgenic mice expressing human cyclin E. *Molecular and Cellular Biology*, 17(1), pp.453-459.

Bueno-de-Mesquita, J.M., Linn, S.C., Keijzer, R., Wesseling, J., Nuyten, D.S.A., Van Krimpen, C., Meijers, C., De Graaf, P.W., Bos, M.M.E.M., Hart, A.A.M. and Rutgers, E.J.T., 2009. Validation of 70-gene prognosis signature in node-negative breast cancer. *Breast cancer research and treatment*, 117(3), pp.483-495.

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

Cheang, M.C., Chia, S.K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P.S., Parker, J.S. and Perou, C.M., 2009. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute*, 101(10), pp.736-750.

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(3-4), pp.273-290.

Cui, J., Antoniou, A.C., Dite, G.S., Southey, M.C., Venter, D.J., Easton, D.F., Giles, G.G., McCredie, M.R. and Hopper, J.L., 2001. After BRCA1 and BRCA2—what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. *The American Journal of Human Genetics*, 68(2), pp.420-431.

Dabbs, D.J., Klein, M.E., Mohsin, S.K., Tubbs, R.R., Shuai, Y. and Bhargava, R., 2011. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the Onco type DX test: an independent quality assurance study. *Journal of Clinical Oncology*, 29(32), pp.4279-4285.

Dowsett, M., Houghton, J., Iden, C., Salter, J., Farndon, J., A'hern, R., Sainsbury, R. and Baum, M., 2006. Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status. *Annals of Oncology*, 17(5), pp.818-826.

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), pp.560-569.

Faneyte, I.F., Schrama, J.G., Peterse, J.L., Remijnse, P.L., Rodenhuis, S. and van De Vijver, M.J., 2003. Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. *British journal of Cancer*, 88(3), pp.406-412.

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), p.R4.

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), p.R4.

Ghosh, S., Sullivan, C.A., Zerkowski, M.P., Molinaro, A.M., Rimm, D.L., Camp, R.L. and Chung, G.G., 2008. High levels of vascular endothelial growth factor and its receptors (VEGFR-1, VEGFR-2, neuropilin-1) are associated with worse outcome in breast cancer. *Human pathology*, 39(12), pp.1835-1843.

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

Gong, M., Dong, W., Shi, Z., Xu, Y., Ni, W. and An, R., 2012. Genetic polymorphisms of GSTM1, GSTT1, and GSTP1 with prostate cancer risk: a meta-analysis of 57 studies. *PLoS One*, 7(11), p.e50587.

Győrffy, 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), pp.725-731.

Haerslev, T., Jacobsen, G.K. and Zedeler, K., 1996. Correlation of growth fraction by Ki-67 and proliferating cell nuclear antigen (PCNA) immunohistochemistry with histopathological parameters and prognosis in primary breast carcinomas. *Breast cancer research and treatment*, 37(2), pp.101-113.

Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M. and Hicks, D.G., 2010. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of pathology & laboratory medicine*, 134(7), pp.e48-e72.

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

Hayes, D.F., Bast, R.C., Desch, C.E., Fritsche, H., Kemeny, N.E., Jessup, J.M., Locker, G.Y., Macdonald, J.S., Mennel, R.G., Norton, L. and Ravdin, P., 1996. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *Journal of the National Cancer Institute*, 88(20), pp.1456-1466.

Hegi, M.E., Diserens, A.C., Godard, S., Dietrich, P.Y., Regli, L., Ostermann, S., Otten, P., Van Melle, G., de Tribolet, N. and Stupp, R., 2004. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. *Clinical cancer research*, 10(6), pp.1871-1874.

Hegi, M.E., Diserens, A.C., Gorlia, T., Hamou, M.F., de Tribolet, N., Weller, M., Kros, J.M., Hainfellner, J.A., Mason, W., Mariani, L. and Bromberg, J.E., 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *New England Journal of Medicine*, 352(10), pp.997-1003.

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

Hsu, K.S. and Kao, H.Y., 2013. Alpha-actinin 4 and tumorigenesis of breast cancer. *Vitamins and hormones*, (93), pp. 323-351.

Hunter, K.W., Crawford, N.P. and Alsarraj, J., 2008. Mechanisms of metastasis. Breast Cancer Research, 10(1), p.S2.

.

Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. and Ménard, S., 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer research*, 65(16), pp.7065-7070.

Jones, R.L., Salter, J., A'Hern, R., Nerurkar, A., Parton, M., Reis-Filho, J.S., Smith, I.E. and Dowsett, M., 2009. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. *Breast cancer research and treatment*, 116(1), pp.53-68.

Karakousis, G., Yang, R. and Xu, X., 2013. Circulating melanoma cells as a predictive biomarker. Journal of Investigative *Dermatology*, 133(6), pp.1460-1462.

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

Lewandowska, M.A., Jóźwicki, W., Starzynski, J. and Kowalewski, J., 2012. Analysis of EGFR mutation frequency and coexistence of KRAS and EGFR mutations using RT-PCR in lung adenocarcinoma: may a clinical and pathological model of a patient's qualification for targeted therapy have an impact on time to obtain genetic results. Pol J *Cardio Thorac Surg*, 9(4), pp.443-51.

Liong, M.L., Lim, C.R., Yang, H., Chao, S., Bong, C.W., Leong, W.S., Das, P.K., Loh, C.S., Lau, B.E., Yu, C.G. and Ooi, E.J.J., 2012. Blood-based biomarkers of aggressive prostate cancer. *PloS one*, 7(9), p.e45802.

Liu, N.Q., Stingl, C., Look, M.P., Smid, M., Braakman, R.B., De Marchi, T., Sieuwerts, A.M., Span, P.N., Sweep, F.C., Linderholm, B.K. and Mangia, A., 2014. Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *Journal of the National Cancer Institute*, 106(2), p.djt376.

Lopez, F., Belloc, F., Lacombe, F., Dumain, P., Reiffers, J., Bernard, P. and Boisseau, M.R., 1991. Modalities of synthesis of Ki67 antigen during the stimulation of lymphocytes. *Cytometry Part A*, 12(1), pp.42-49.

Lowery, A.J., Miller, N., Devaney, A., McNeill, R.E., Davoren, P.A., Lemetre, C., Benes, V., Schmidt, S., Blake, J., Ball, G. and Kerin, M.J., 2009. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Research*, 11(3), p.R27.

Mass, R.D., Press, M.F., Anderson, S., Cobleigh, M.A., Vogel, C.L., Dybdal, N., Leiberman, G. and Slamon, D.J., 2005. Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clinical breast cancer*, 6(3), pp.240-246.

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), pp.1-7.

Nalejska E, Mączyńska E, Lewandowska MA. Prognostic and Predictive Biomarkers: Tools in Personalized Oncology. *Molecular Diagnosis & Therapy*. 2014;18(3):273-284. doi:10.1007/s40291-013-0077-9.

Nguyen, B., Cusumano, P.G., Deck, K., Kerlin, D., Garcia, A.A., Barone, J.L., Rivera, E., Yao, K., de Snoo, F.A., van den Akker, J. and Stork-Sloots, L., 2012. Comparison of molecular subtyping with BluePrint, MammaPrint, and TargetPrint to local clinical subtyping in breast cancer patients. *Annals of surgical oncology*, 19(10), pp.3257-3263.

Nicholson, R.I., McClelland, R.A., Finlay, P., Eaton, C.L., Gullick, W.J., Dixon, A.R., Robertson, J.F.R., Blamey, R.W. and Ellis, I.O., 1993. Relationship between EGF-R, c-erbB-2 protein expression and Ki67 immunostaining in breast cancer and hormone sensitivity. *European journal of cancer*, 29(7), pp.1018-1023.

Ormandy, C.J., Musgrove, E.A., Hui, R., Daly, R.J. and Sutherland, R.L., 2003. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast cancer research and treatment*, 78(3), pp.323-335.

Paduch, R. and Klatka, J., 2003. Markery nowotworowe. *Onkol*. Pol, 2, pp.77-82.

PANCZYK, M. and Mirowski, M., 2008. Farmakogenetyka-znaczenie w chemioterapii raka jelita grubego. *Nowotwory Journal of Oncology*, 58(1), pp.62-69.

Parise, C.A. and Caggiano, V., 2014. Breast cancer survival defined by the ER/PR/HER2 subtypes and a surrogate classification according to tumor grade and immunohistochemical biomarkers. *Journal of cancer epidemiology*, 2014.

Qi, J., Wang, J., Katayama, H., Sen, S. and Liu, S.M., 2013. Circulating microRNAs (cmiRNAs) as novel potential biomarkers for hepatocellular carcinoma. *Neoplasma*, 60(2), p.135.

Renganathan, Vinaitheerthan 2017. Text Mining in Biomedical Domain with Emphasis on Document Clustering. *Healthcare Informatics Research*. 23: 141–146 – via.

Ringnér, M., Fredlund, E., Häkkinen, J., Borg, Å. and Staaf, J., 2011. GOBO: gene expression-based outcome for breast cancer online. *PloS one*, 6(3), p.e17911.

Romagnoli, M., Belguise, K., Yu, Z., Wang, X., Landesman-Bollag, E., Seldin, D.C., Chalbos, D., Barillé-Nion, S., Jézéquel, 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), pp.6268-6278.

Rudolph, P., Olsson, H., Bonatz, G., Ratjen, V., Bolte, H., Baldetorp, B., Fernö, M., Parwaresch, R. and Alm, P., 1999. Correlation between p53, c-erbB-2, and topoisomerase II alpha expression, DNA ploidy, hormonal receptor status and proliferation in 356 node-negative breast carcinomas: prognostic implications. *Journal of pathology*, 187(2), pp.16-207.

Rydén, L., Stendahl, M., Jonsson, H., Emdin, S., Bengtsson, N.O. and Landberg, G., 2005. Tumor-specific VEGF-A and VEGFR2 in postmenopausal breast cancer patients with long-term follow-up. Implication of a link between VEGF pathway and tamoxifen response. *Breast cancer research and treatment*, 89(2), pp.135-143.

Samzadeh, M., Hasanzad, M., Jamaldini, S.H., Haghdoost, A.A., Afshari, M. and Ziaee, S.A.M., 2012. Association of G/A polymorphism, rs266882, in AREI region of the prostate-specific antigen gene with prostate cancer risk and clinicopathological features. *Urology journal*, 9(4), p.691.

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), pp.72-77.

Smith, M.L. and Seo, Y.R., 1999. Sensitivity of cyclin E-overexpressing cells to cisplatin/taxol combinations. *Anticancer research*, 20(4), pp.2537-2539.

Synowiec, EMA., Krupa, R., Woz´niak, K. 2010. Gene's polymorphism of DNA double-stranded repair in breast cancer. *Post Biol Kom*.37(2):433–8.

van den Bent, M.J., Hegi, M.E. and Stupp, R., 2006. Recent developments in the use of chemotherapy in brain tumours. *European Journal of Cancer*, 42(5), pp.582-588.

Van't Veer, L.J., Dai, H., Van De Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. and Schreiber, G.J.,

2002. Gene expression profiling predicts clinical outcome of breast cancer. *nature*, 415(6871), pp.530-536.

Vasiljević, N., Ahmad, A.S., Beesley, C., Thorat, M.A., Fisher, G., Berney, D.M., Møller, H., Yu, Y., Lu, Y.J., Cuzick, J. and Foster, C.S., 2013. Association between DNA methylation of HSPB1 and death in low Gleason score prostate cancer. *Prostate cancer and prostatic diseases*, 16(1), pp.35-40.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. and Prueitt, R.L., 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National academy of Sciences of the United States of America*, 103(7), pp.2257-2261.

Voon, P.J. and Kong, H.L., 2011. Tumour genetics and genomics to personalise cancer treatment. *Annals of the Academy of Medicine-Singapore*, 40(8), p.362.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I. and Kerr, D., 2009. Genetic prognostic and predictive markers in colorectal cancer. Nature Reviews Cancer, 9(7), pp.489-499.

Wang, F.B., Yang, X.Q., Yang, S., Wang, B.C., Feng, M.H. and Tu, J.C., 2011. A higher number of circulating tumor cells (CTC) in peripheral blood indicates poor prognosis in prostate cancer patients--a meta-analysis. *Asian Pac J Cancer Prev*, 12(10), pp.2629-35.

Wingate, H., Puskas, A., Duong, M., Bui, T., Richardson, D., Liu, Y., Tucker, S.L., Van Pelt, C., Meijer, L., Hunt, K. and Keyomarsi, K., 2009. Low molecular weight cyclin E is specific in breast cancer and is associated with mechanisms of tumor progression. *Cell Cycle*, 8(7), pp.1062-1068.

Wotschofsky, Z., Busch, J., Jung, M., Kempkensteffen, C., Weikert, S., Schaser, K.D., Melcher, I., Kilic, E., Miller, K., Kristiansen, G. and Erbersdobler, A., 2013. Diagnostic and prognostic potential of differentially expressed miRNAs between metastatic and non-metastatic renal cell carcinoma at the time of nephrectomy. *Clinica chimica acta*, 416, pp.5-10.

Xu, J., Zheng, S.L., Isaacs, S.D., Wiley, K.E., Wiklund, F., Sun, J., Kader, A.K., Li, G., Purcell, L.D., Kim, S.T. and Hsu, F.C., 2010. Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proceedings of the National Academy of Sciences*, 107(5), pp.2136-2140.

Xu, Y., Sun, Y., Yao, L., Shi, L., Wu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T. and Lin, B., 2008. Association between CYP2D6* 10 genotype and survival of breast cancer patients receiving tamoxifen treatment. *Annals of oncology*, 19(8), pp.1423-1429.

CHAPTER 4

4. Expression analysis of microRNA and target genes using *in silico* approaches

4.1 Introduction

MicroRNAs influence nearly all types of cellular pathways including oncogenesis and function in multiple-to-multiple relationships with their target genes (Zhang *et al.*, 2006). Moreover, while a single microRNA can target multiple mRNAs, a single mRNA can be targeted by multiple microRNAs (Hsu *et al.*, 2014). It is suggested that the final protein output is defined by microRNA expression levels as well as the availability of mRNA targets, thus, change in the expression level of a particular microRNA leads to severe pathological conditions (Wang & Wang, 2006; Abdellatif, 2012; Sheedy, 2015). Studying these microRNA expression profiles will provide new methods for tumors classification and cancer diagnosis as described by Berindan-Neagoe *et al.*, (2014); & Lu *et al.* (2005).

With the advancement of the microarray and the RNA sequencing technologies, an exponential flow of gene expression data has emerged, unveiling gene behaviour under varying biological conditions. Using microarray expression profiling, identification of up- and down-regulated genes and microRNAs in cancer cells become highly feasible. Differential analysis of microRNA gene expression levels in

normal versus cancer cells/tissues provides insights about the functioning of microRNAs, which can be utilized in cancer diagnostics and therapeutics. The preliminary detection of differentially expressed (DE) microRNAs from large-scale microRNA expression profiling data can be efficiently done using robust and efficient statistical models of bioinformatics (Dressman *et al*., 2003; Subramanian *et al*., 2004; Glanzer & Eberwine, 2004). Using bioinformatics tools in this chapter, comparative analysis of gene expression patterns of cancer cells with normal cells or other subtypes of the cancer were identified and clustered. Additionally, biological functions of the selected signature genes were studied to develop cancer biomarkers for early diagnosis and prognosis. Overall integrating the gene and microRNA expression data with bioinformatics tools will help to identify microRNAs as novel biomarkers for cancer diagnosis and prognosis as well as to understand their role in cancer progression and metastasis (Lu *et al*., 2005; Edelman *et al*., 2010).

4.2 MicroRNA expression database for tissues

4.2.1 miRmine (Human miRNA expression database)

The miRmine database was developed to study microRNA expression profiles from publicly available human microRNA sequencing data. This database serves as an integrated user-interactive web-resource for the relative abundance of different human microRNAs expression in multiple tissue and cell-line. In miRmine, single or multiple (comma separated) human microRNAs can be searched using standard accession IDs (e.g., hsa-miR-21-5p). Furthermore, the search can also be done from

any sub-part of an accession ID (e.g., miR-21 or 21) using an auto-completion option where the corresponding entries will be automatically generated from the database (Panwar *et al*., 2017).

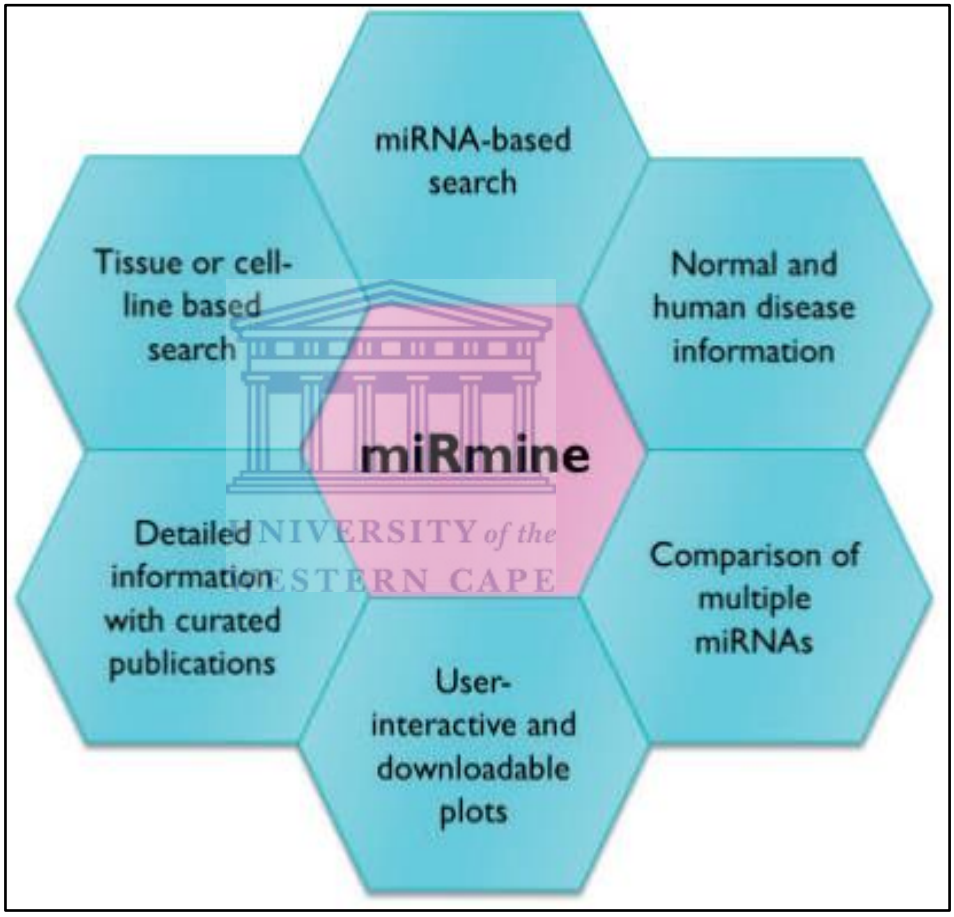


Figure 4.1: Different applications of the miRmine database (adapted from Panwar *et al*., 2017).

4.2.2 MicroRNA target genes expression databases for tissues

4.2.2.1 Oncomine

ONCOMINE is a cancer microarray database and web-based data-mining platform made up of 65 gene expression datasets, and is available at www.oncomine.org. These datasets comprise nearly 48 million gene expression measurements collected from over 4700 microarray experiments. ONCOMINE was designed using three levels. Level 1 comprises an Oracle 8i database to store microarray data, other biological data and statistics. Level 2 is constructed from Python (www.python.org) which requires handling application logic and core functionality, while level 3 is developed using ZOPE (www.zope.org) (Rhodes *et al*., 2004).

While Oncomine was developed to facilitate ascertaining genome-wide expression analysis, it also allows the comparative differential expression of major types of cancer with respect to normal tissues. Furthermore, comparison of gene expression between cancer subtypes and clinical- and pathology-based analyses can also be explored using this tool. (Rhodes *et al*., 2004). Oncomine outcomes can be visualized for single or multiple genes in selected or grouped analyses. Additionally, specific gene annotations like secreted, kinase, membrane, etc. can be used (Rhodes *et al*., 2004). Oncomine can integrate information from other bioinformatics resources and other databases including Swiss-Prot, Locus Link (Pruitt & Maglott, 2001); Unigene, Human Protein Reference Database (HPRD) (Navarro *et al*., 2003); SOURCE (Diehn

et al., 2003); Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) and Biocarta (Rhodes *et al*., 2004).

4.2.2 FIREBROWSE

FIREBROWSE is a web browser for easy download and analysis of TCGA data in a simple way. The Cancer Genome Atlas or TCGA is a public systematic cancer genomics project launched by the National Collaborative Research Infrastructure (NCI) and the National Human Genome Research Institute (NHGRI) (http://cancergenome.nih.gov/). TCGA contains molecular profiles of specific tumor types characterized by using advanced technologies that evaluate the sequence of the exome, SNP arrays, DNA methylation, mRNA expression and sequence, microRNA expression and transcript splice variation (Weinstein *et al*., 2013; Kandoth *et al*., 2013). FIREBROWSE is backed by a powerful computational infrastructure, application programming interface (API), graphical tools and online reports, which can be browsed at URL http://firebrowse.org. It is one of the integrative characterized open cancer datasets in the world with over 80K sample aliquots from more than 11,000+ cancer patients, spanning 38 unique disease cohorts. Using FIREBROWSE**,** TCGA data and analysis in cBioPortal—expression, mutation, copy number, significance analysis, and other operations can be loaded directly (FIREBROWSE , 2016).

4.3 Aim and objectives

The aim of this chapter is to analyze the expression values the five novel microRNAs and their associated target genes (nine) as biomarkers in BC using different bioinformatics tools. The study was conducted using the following objectives.

i. Expression analysis of the microRNA in tissue using miRmine.
ii. Expression analysis of the target genes regulated by these microRNAs individually in different stages of BC using Oncomine.
iii. Expression analysis of the target genes regulated by these microRNAs individually in different cancer using FIREBROWSE.

4.4 Materials and methods.

The expression value of the five novel microRNAs and their associated nine target genes as biomarkers in BC was analyzed using the bioinformatics tools, viz. miRmine, Oncomine and FIREBROWSE.

4.4.1 Expression analysis of the microRNAs in tissue using miRmine

The high-quality experimental data from 16 different types of human tissues and biofluids were analysed viz. all tissues, bladder, blood, brain, breast, liver, lung, nasopharynx, pancreas, placenta, plasma, saliva, semen, serum, sperm and testis. All microRNAs were entered into the miRmine search box together using the microRNA nomenclature, ‘hsa-miR-’. ‘Human’ was selected as the species for expression

analysis for microRNAs. Thereafter, the ‘Tissue’ option was chosen followed by the ‘Run’ option.

4.4.2 Expression analysis of the microRNA target genes in different stages of breast cancer using Oncomine

In the Oncomine database (www.oncomine.org), key words and phrases that allowed extraction of genes specifically for BC tissue such as <differential expression>, <breast cancer>, <cancer vs. cancer analysis>, <TCGA breast 2 Dataset>, and <Grouped by Breast Adenocarcinoma Type> were used as input queries. The database of genes and *Homo sapiens* genes were selected. The results from the ONCOMINE database were saved in a separate file and each gene was categorized according to BC type.

4.4.3 Expression analysis of the microRNA target genes in different cancers using FIREBROWSE

The FIREBROWSE homepage is accesible at http://firebrowse.org/. The gene names were individually entered into the *search box*, and then clicked on the *view expression profile.* For Graphical Tools *filter* was selected by clicking on the *on* position, then *all cohorts* were selected and the *submit* option pressed.

4.5 Results

4.5.1 Expression of microRNAs in different tissues using miRmine

Figure 4.2 illustrates the expression of five novel microRNAs in different tissues suggesting their tissue-specific availability. Overall, four of the five microRNAs, (miR1, miR2, miR4, and miR5) were expressed in all the organs. All five microRNAs were expressed in sperm,whereas only four were expressed in the breast tissue. Three microRNAs (MiR1,MiR2 and MiR4) were expressed in bladder, brain, nasopharynx, pancreas, and semen. Liver, placenta, plasma, and testis showed expression of two microRNAs, while lung and saliva displayed expression of a single microRNA.

MiR1 was expressed inthe bladder, brain, breast, nasopharynx, pancreas, and sperm, with the highest expression in observed breast tissue. MiR2 was expressed in almost all tissues except lung tissue, while the highest expression was observed in sperm. MiR3 was found to be expressed in sperm alone, whilst MiR4 was reported to be expressed in all tissues except saliva and serum. The highest expression of MiR4 was observed in breast tissue. MiR5 was observed expressed in breast tissue and semen, with the highest expression in breast tissue.

Table 4.1 demonstrates the expression values of all the microRNAs in breast tissue, a serum of cancer and normal controls. While miR1 was found to be expressed in

serum samples of cancer patients and normal controls, it was reported to be absent in tissue samples. In the case of serum, the expression level of MiR1 was elevated by 51.6 RPM units in the diseased condition. While for miR2 the expression was observed in the serum of cancer patients and normal control samples, but was only expressed in normal tissue. In the case of cancer patient's serum samples, it was found to be elevated by 149.4 RPM units when compared with a normal control sample. MiR3 expression was altogether absent in all the breast tissue samples, while MiR4 expression was down-regulated to 88.3 RPM units in cancer patients' serum in comparison with a normal control sample. No noticeable change was observed in tissue samples. Finally, miR5 was absent in serum samples and to be marginally up-regulated to 17.4 RPM units in BC tissue in comparison to a normal control sample.

Table 4.1: Expression values (RPM) of miR1, miR2, miR3, miR4 and miR5 in breast tissues.

Show 10 entries Search:

Experiment ID	Sample ID	Tissue	Description	Disease	Sex	PubMed ID		miiR3	miiR1	miiR4	miiR5	miiR2
SRX513283	SRS590232	Breast	Tumor serum	Breast cancer		24904649	0	0	190.1	142.6	0	380.3
SRX513284	SRS590233	Breast	Tumor tissue	Breast cancer		24904649	0	0	0	348.1	87	0
SRX513285	SRS590234	Breast	Serum from healthy woman control	Normal		24904649	0	0	138.5	230.9	0	230.9
SRX513286	SRS590235	Breast	Normal tissue	Normal		24904649	0	0	0	347.8	69.6	69.6

Showing 1 to 4 of 4 entries First Previous 1 Next Last

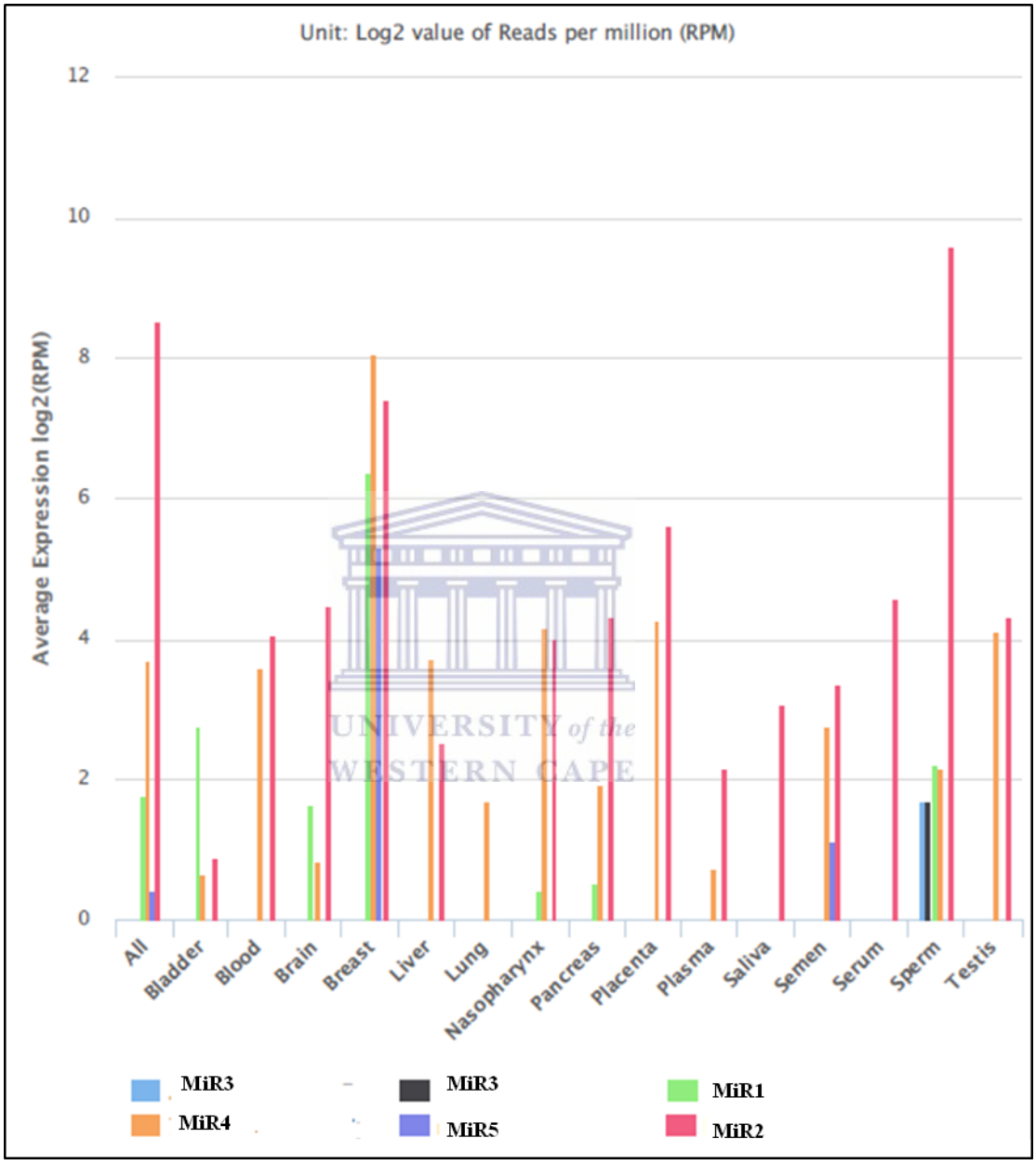


Figure 4.2: Expression of microRNAs in different human tissues.

4.5.2 Expression of the microRNA target genes in different breast adenocarcinoma types

From previous results, five microRNAs were regulate a total of nine gene targets. All these target genes were further studied for their expression levels in multiple breast adenocarcinoma types. MiR1 was regulated target gene *HIP1*, whose expression in different breast adenocarcinomas is represented in Figure 4.3 Overall, the expression of *HIP1* gene was partially over-expressed in all breast adenocarcinomas except for invasive cribriform breast carcinoma. Maximum up-regulation of *HIP1* gene was observed in papillary breast carcinoma. Uniform up-regulation in all the samples were obtained in medullary and papillary breast carcinoma, while ductal and lobular, either mixed or alone, and mucinous breast carcinoma samples displayed differential expression.

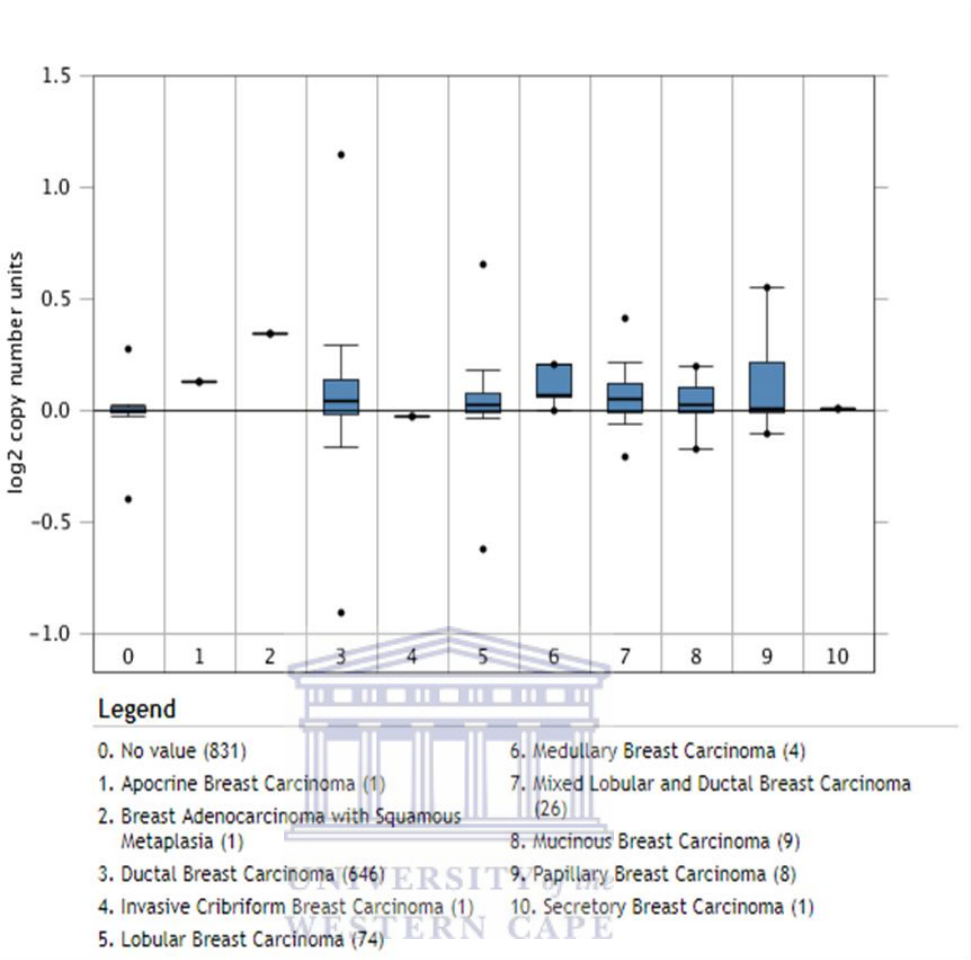


Figure 4.3: Target gene for MiR1 - *HIP1* copy number in TCGA breast 2 grouped by breast adenocarcinoma type

In Figure 4.4, the expression of *SYT9*, the target gene of MiR2 in different types of breast adenocarcinomas, is represented. While the expression of this target gene was found to be completely down-regulated in breast adenocarcinoma with squamous metaplasia and papillary BC, it is up-regulated in apocrine and invasive cribriform BC. In remaining breast carcinoma types, it is found to be differentially expressed

with the few samples tested showing elevated expression, whereas other samples tested show reduced expression

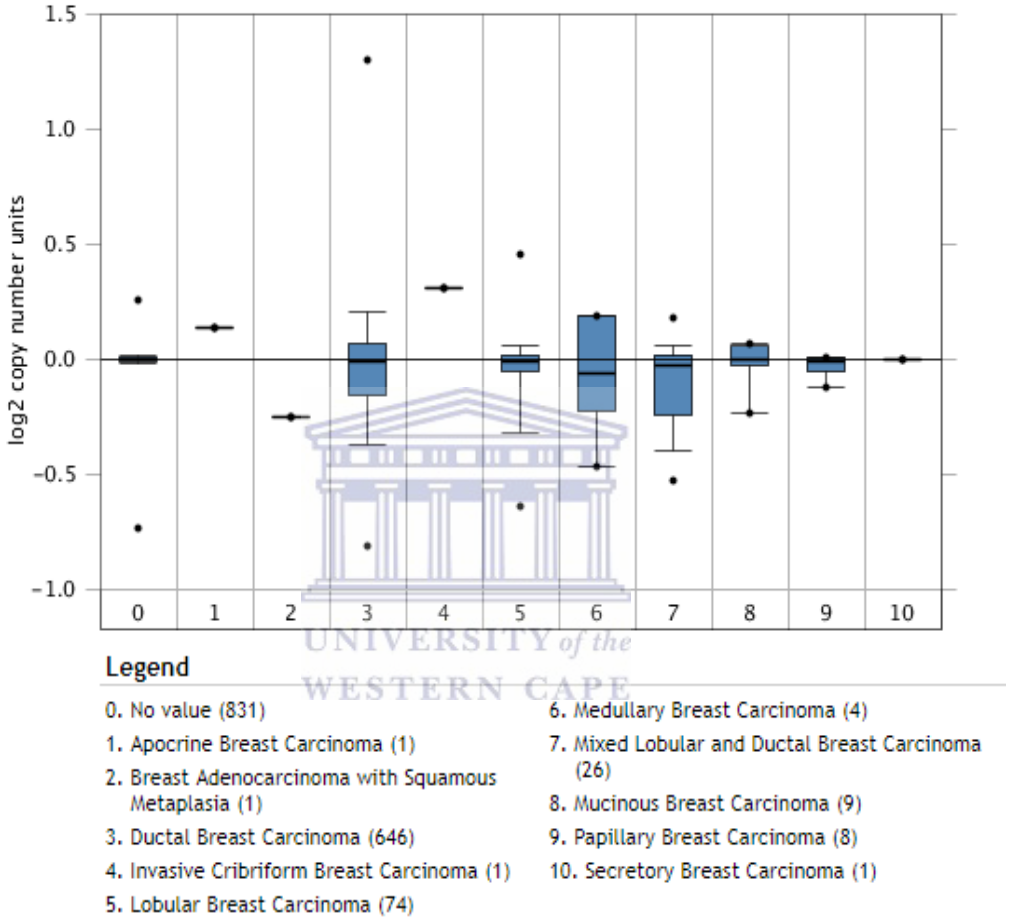


Figure 4.4: Target gene for MiR2 – *SYT9* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

MiR3 was regulated two target genes, (*TEMD10* and *AP1G1),* whose expression in multiple breast adenocarcinomas were investigated. As demonstrated in Figure 4.5,

the expression of *TEMD10* was down-regulated in apocrine, squamous metaplasia type, invasive cribriform, medullary and papillary breast carcinoma. Ductal and lobular breast carcinoma samples revealed differential expression with few samples showing elevated expression. While a large number of samples showed down-regulation, an opposite trend was seen in mixed lobular and ductal as well as mucinous breast carcinoma samples. Similarly, the expression of *AP1G1*,(another target gene of MiR3), was overall down-regulated in all types of breast adenocarcinoma except for mucinous breast carcinoma, which showed partial up-regulation (Figure 4.6).

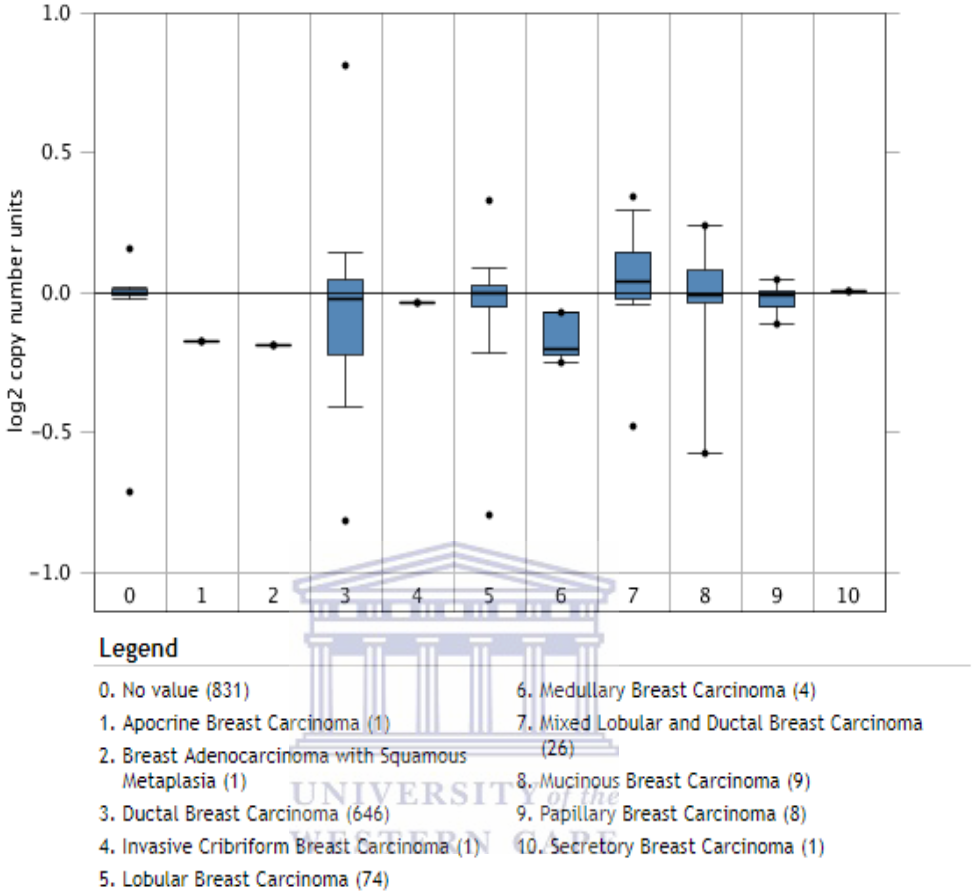


Figure 4.5: Target gene for MiR3 – *TEMD10* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

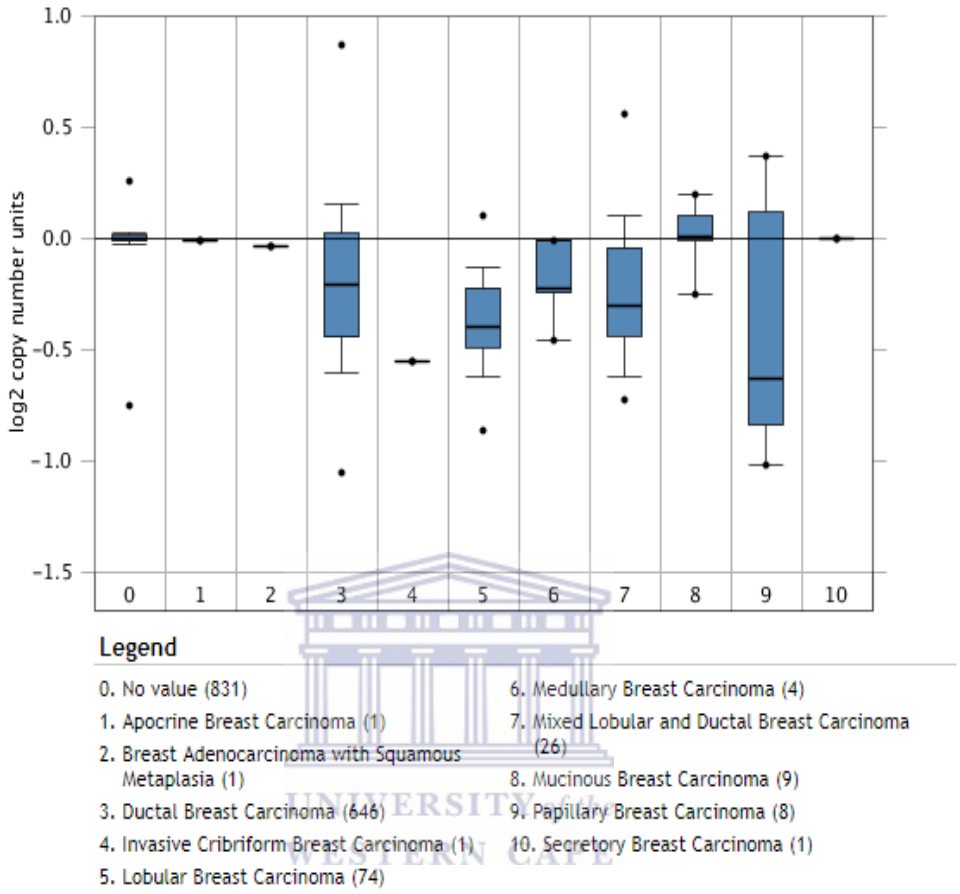


Figure 4.6: Target gene for MiR3 – *AP1G1* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

Like MiR3, MiR4 also regulates multiple target genes, (*CD2AP*, *ANP32E*, and *VEGFA)*, as demonstrated in Figures 4.7- 4.9. Both *CD2AP* and *VEGFA* were marginally down-regulated in apocrine, squamous metaplasia, invasive cribriform and secretory breast carcinoma, while its expression is elevated in medullary and mucinous breast carcinoma. In the remaining BC types, *CD2AP* and *VEGFA* showed

differential expression, with few samples showing up-regulation, while other samples showed expression of down-regulation. On the contrary, *ANP32E* was overall up-regulated in all types of breast adenocarcinoma.

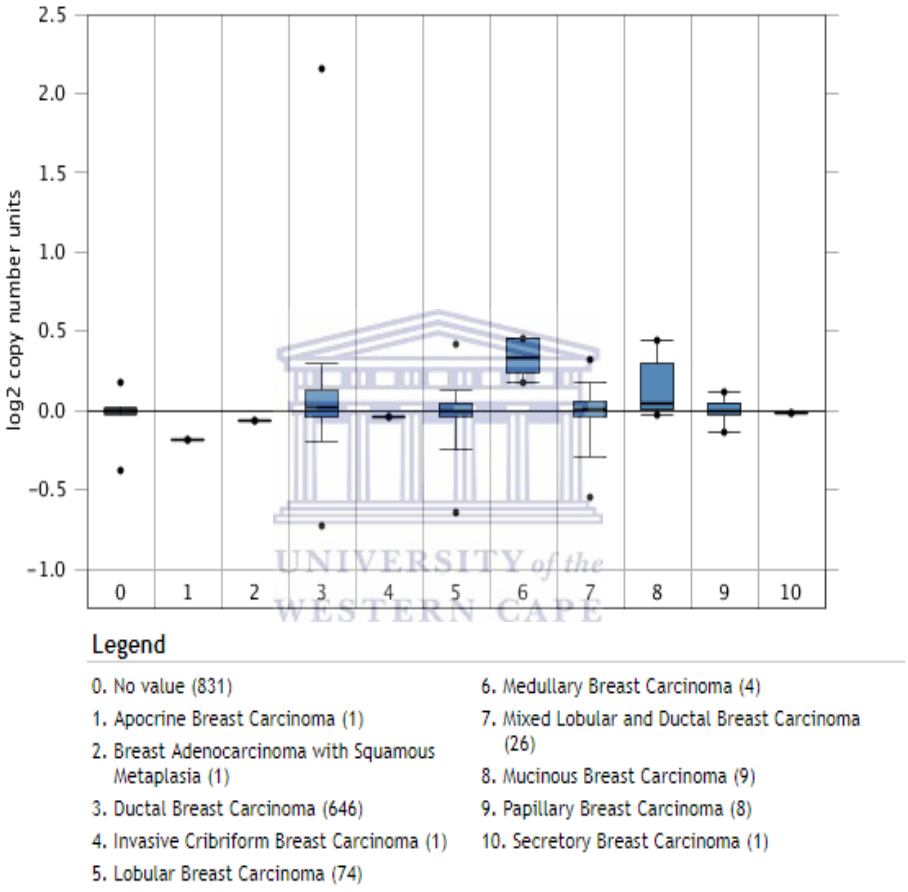


Figure 4.7: Target gene for MiR4 – CD2AP copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

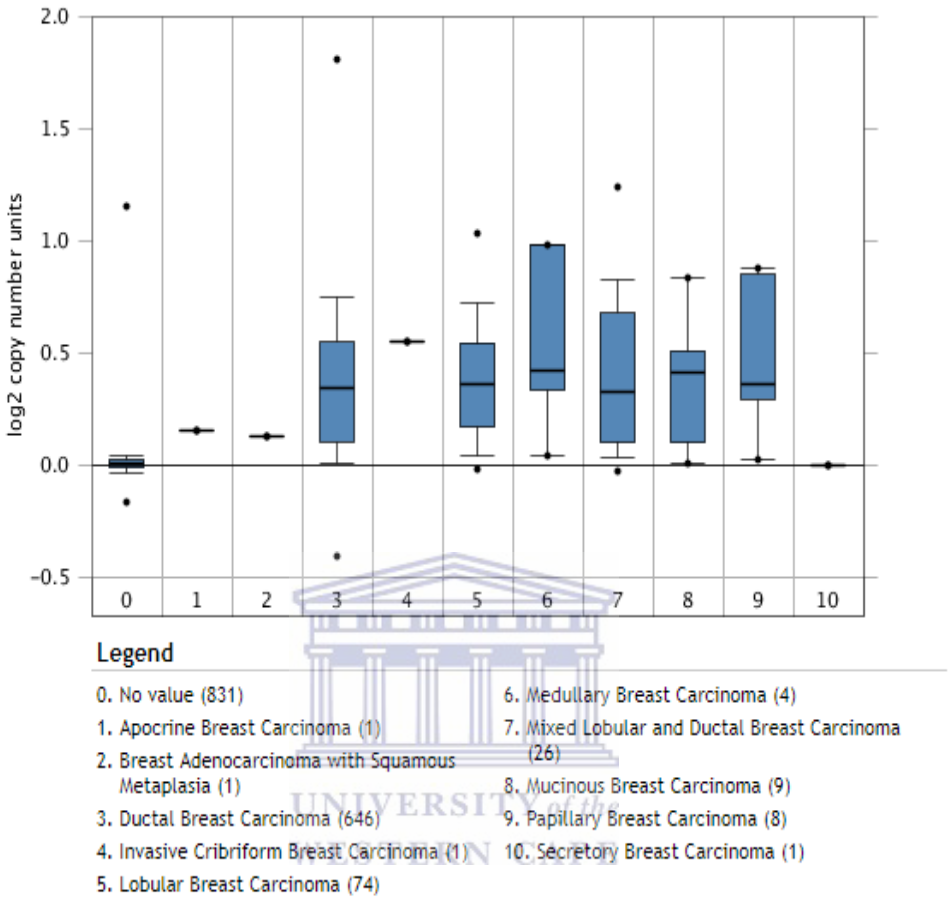


Figure 4.8: Target gene for MiR4 – *ANP32E* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

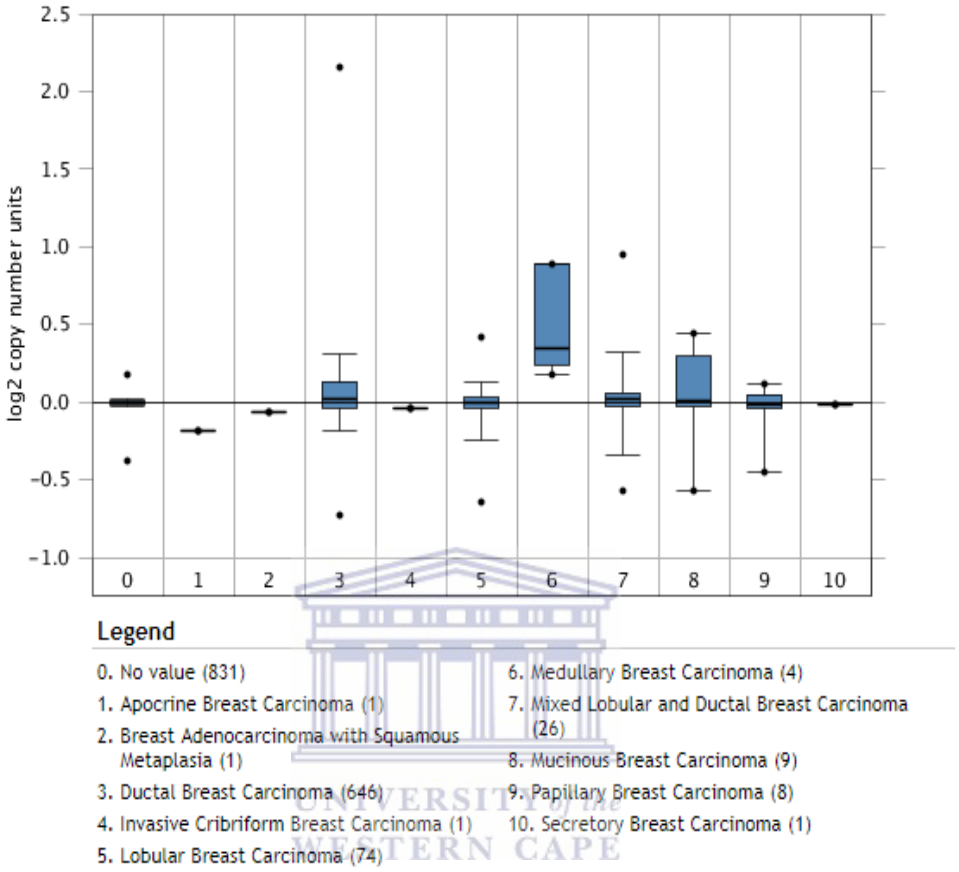


Figure 4.9: Target gene for MiR4 – *VEGFA* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

Finally, MiR5 was targeted two target genes (*GOPC* and *ACTN4).* As demonstrated in Figure 4.10, *GOPC* was mostly down-regulated in all types of breast adenocarcinoma except for apocrine and medullary breast carcinoma. Similarly, *ACTN4* was up-regulated in apocrine and medullary breast carcinoma and marginally down-regulated in squamous metaplasia, invasive cribriform and secretory breast

carcinoma. In the remaining BC types, *ACTN4* showed differential expression, with few samples showing up-regulated, while other samples showed down-regulated expression (Figure 4.11).

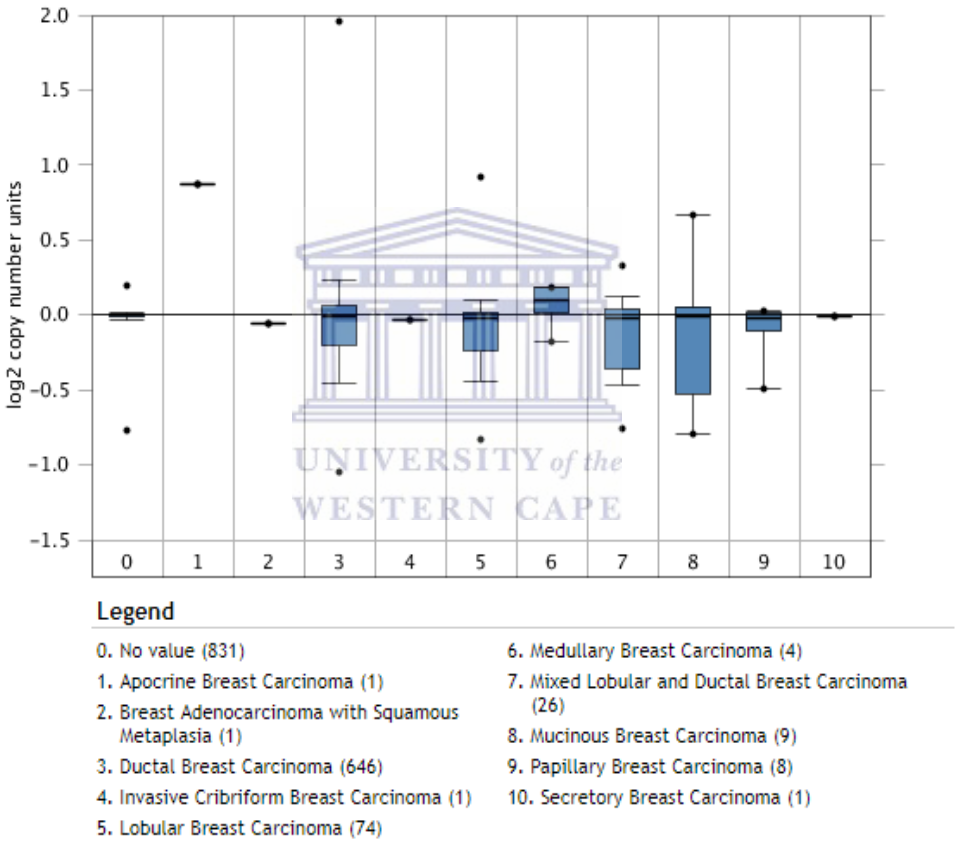


Figure 4.10: Target gene for MiR5 – *GOPC* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

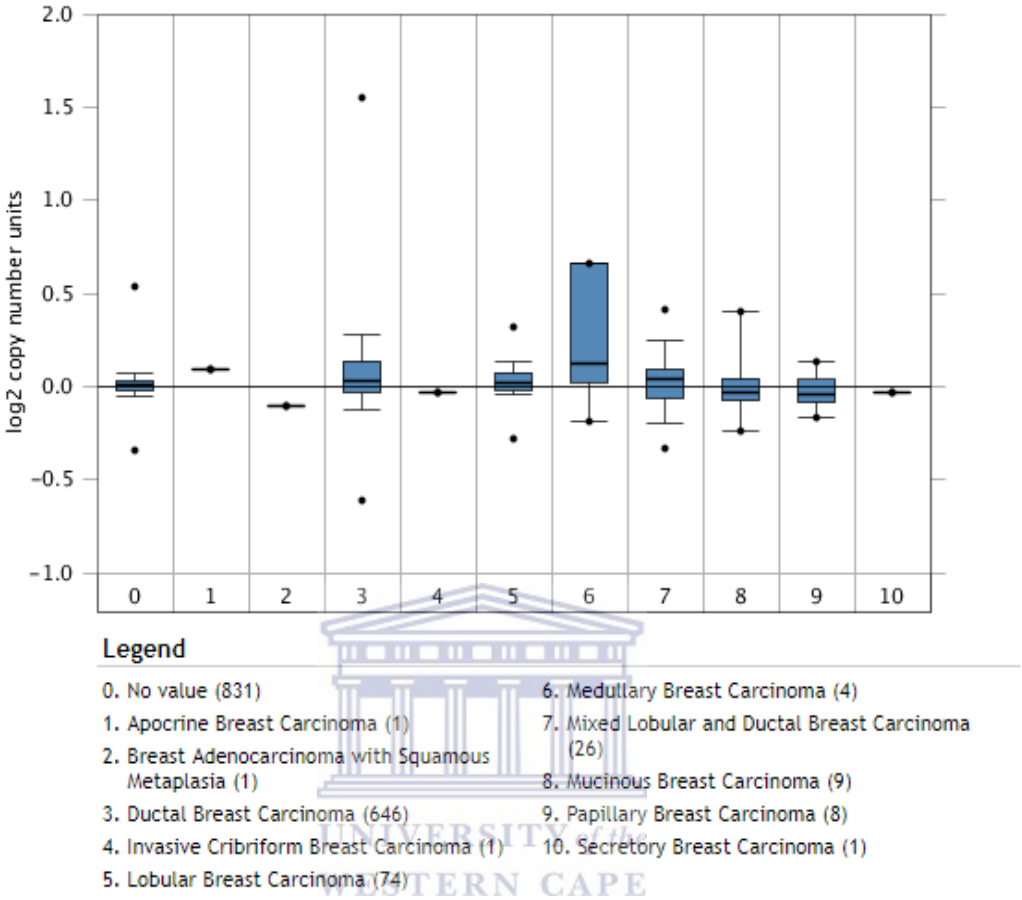


Figure 4.11: Target gene for MiR5 – *ACTN4* copy number in TCGA breast 2 grouped by breast adenocarcinoma type.

4.5.3 Expression of the microRNA target genes in different cancerous tissues

In addition to the previous analysis, expression of the nine gene targets of the five novel BC microRNAs was evaluated in different tissues. Expression of target gene HIP1, regulated by MiR1, showed mixed response as demonstrated in Figure 4.12. In comparison to normal tissues, the *HIP1* gene was significantly over-expressed in

head and neck squamous cell carcinoma (HNSC), liver hepatocellular carcinoma, (LIHC), stomach adenocarcinoma (STAD), and stomach and esophageal carcinoma (STES). In esophageal carcinoma (ESCA) and breast invasive carcinoma (BRCA) cancer tissues, the *HIP1* gene was only partially expressed. On the contrary, its expression was considerably down-regulated in lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and uterine corpus endometrial carcinoma (UCEC). In bladder urothelial carcinoma (BLCA), pan-kidney cohort (KICH+KIRC+KIRP (KIPAN), and kidney renal clear cell carcinoma (KIRC) this gene was partially reduced in comparison to normal tissues. In the remaining tissues, its expression was either unaffected or inappropriately reported.

Figure 4.13 represents the expression of *SYT9*, the target gene of MiR2, in different cancer tissues. In comparison to the normal tissues, *SYT9* is partially expressed in BRCA. While *SYT9* was significattly mostly reduced in cancer tissues, its expression was noticeably down-regulated in BLCA, ESCA, LIHC, LUAD, and UCEC in comparison to normal tissues. In KIPAN and KIRC cancer tissues, its expression was partially reduced.

As shown in Figure 4.14, the target gene of MiR3, *TEMD10*, was over-expressed in BLCA, ESCA, HNSC, LIHC, LUAD, LUSC, STAD, STES, and UCEC cancer tissues in comparison to normal tissues. In BRCA this was partially elevated when

compared with normal controls. On the contrary, *AP1G1* another target gene of MiR3, showed differential expression. While in BLCA, LIHC, and STAD tissues it was found to be up-regulated, it was remarkably down-regulated in KIPAN and KIRC. Furthermore, it was also found to be reduced in LUAD and LUSC in comparison to normal tissues (Figure 4.15).

In previous results, MiR4 was found to express three target genes – *CD2AP*, *ANP32E,* and *VEGFA*. According to in Figure 4.16, the expression level of *VEGFA*, was significantly up-regulated in KIPAN and KIRC cancer tissues, but only partially elevated in BRCA, ESCA, HNSC, LIHC, STAD, and STES when compared normal controls. Similarly, another target gene of MiR4, ANP32E expression was up-regulated in most tissues. While in ESCA, HNSC, STAD, and STES significant increase in elevation of expression was observed. Howevere, only partial up-regulation was observed in BLCA, BRCA, LIHC, LUAD, and LUSC in when compared to normal controls (Figure 4.17). Finally, *CD2AP* showed a mixed response, while in cancer tissues like BLCA, LIHC, LUAD, STAD, STES, and UCEC it showed up-regulated expression. However, in KIPAN and KIRC it was considerably down-regulated when compared to normal controls (Figure 4.18).

Earlier, *GOPC* and *ACTN4* were obtained as the target genes of novel MiR5, whose expression in different cancer tissues are represented in Figures 4.19 and 4.20, respectively. While *GOPC* expression was shown to be up-regulated in HNSC and

LIHC cancer tissues, *ACTN4* expression was elevated in ESCA and UCEC. Moreover, in BRCA, *GOPC* was partially reduced compared to the normal tissues, while *ACTN4* expression was the same in the normal and cancer tissues in BRCA.

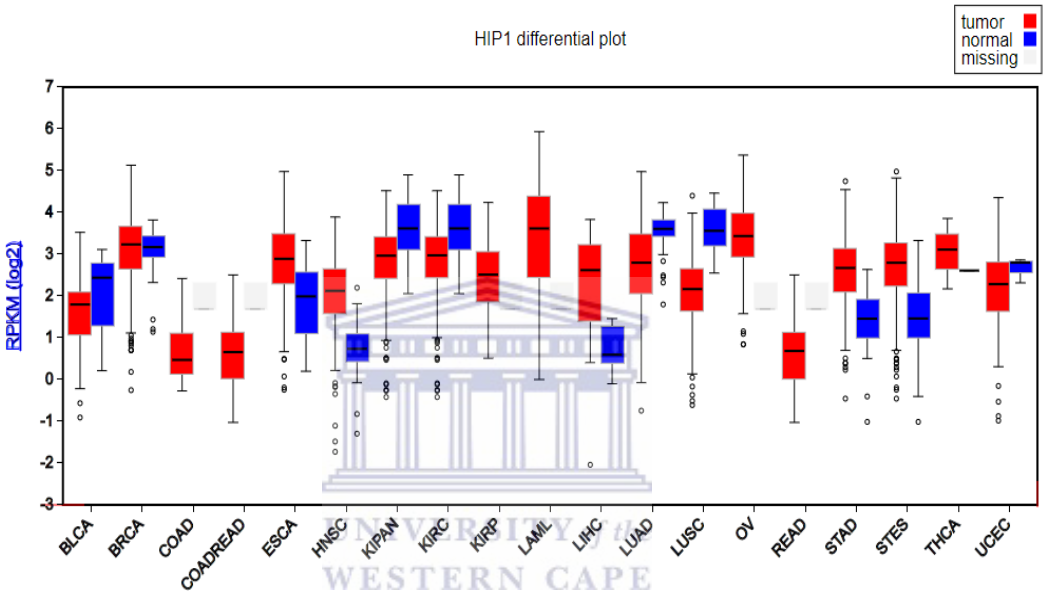


Figure 4.12: Target gene for MiR1 - *HIP1* copy number in different cancer tissues.

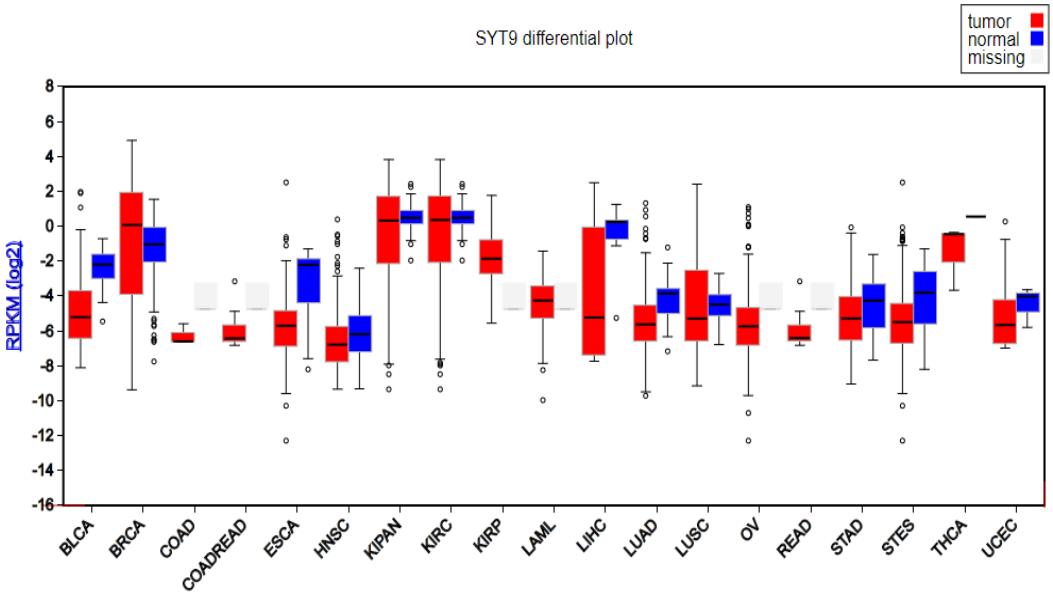


Figure 4.13: Target gene for MiR2 – *SYT9* copy number in different cancer tissues.

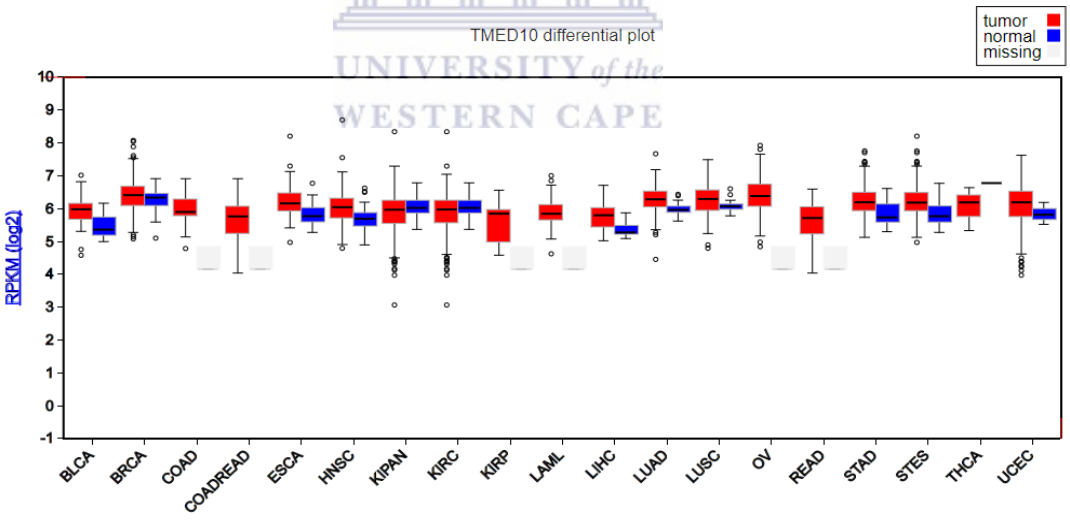


Figure 4.14: Target gene for MiR3 – *TEMD10* copy number in different cancer tissues.

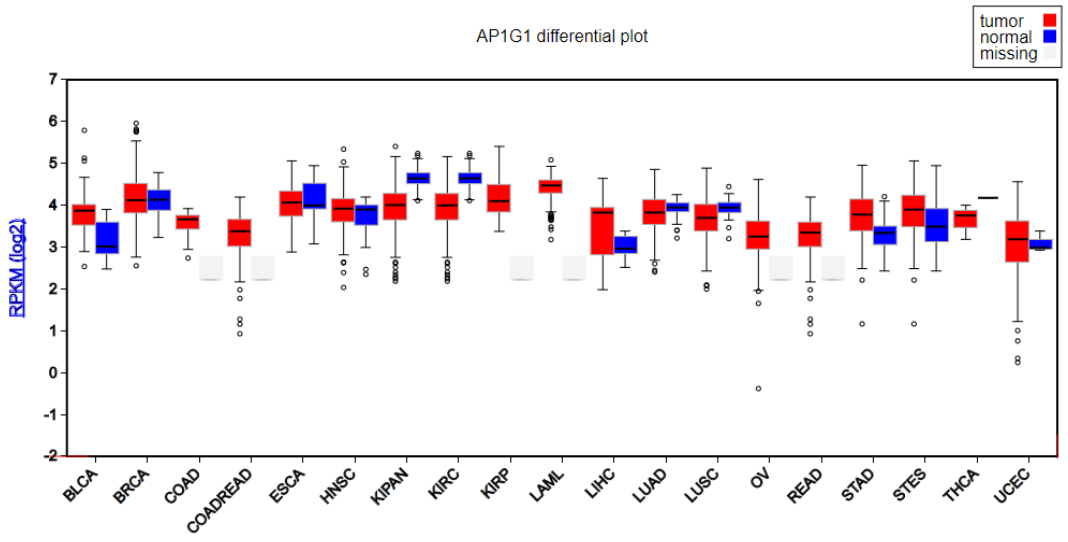


Figure 4.15: Target gene for MiR3 – *AP1G1* copy number in different cancer tissues.

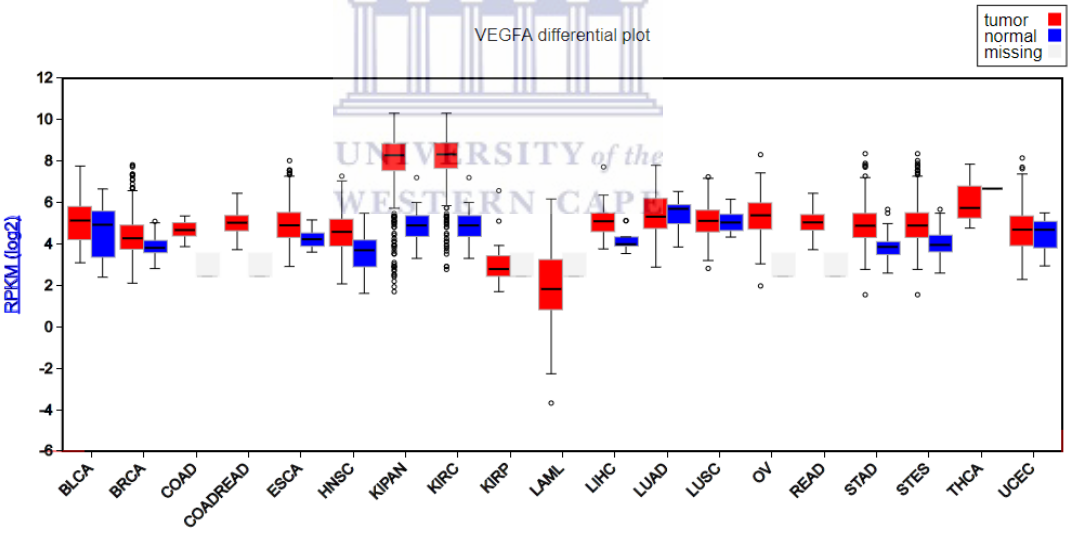


Figure 4.16: Target gene for MiR4 – *CD2AP* copy number in different cancer tissues.

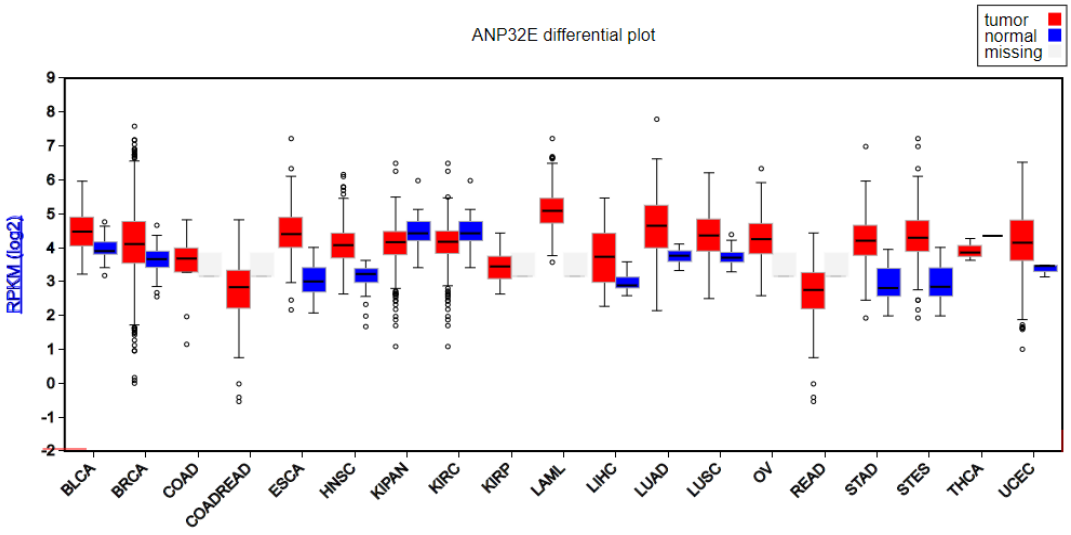


Figure 4.17: Target gene for MiR4 – *ANP32E* copy number in different cancer tissues.

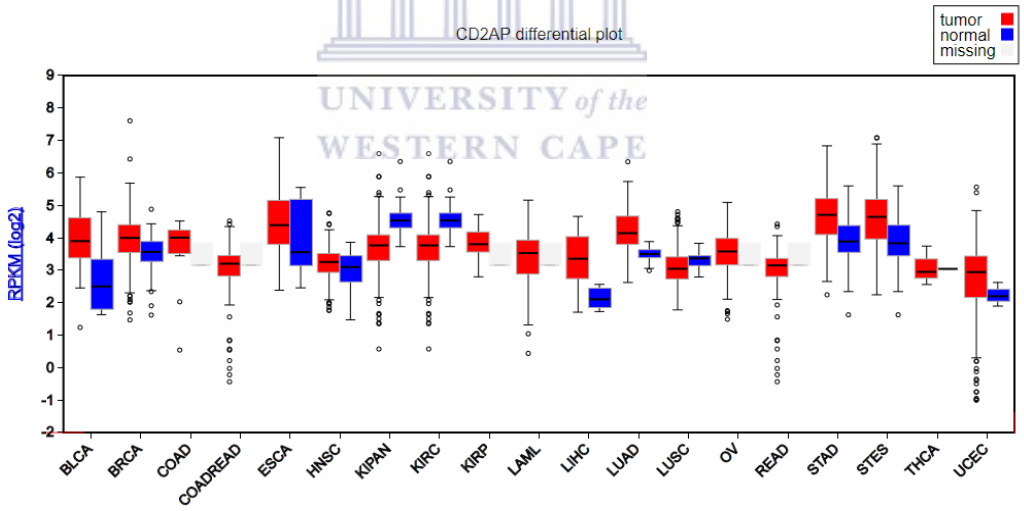


Figure 4.18: Target gene for MiR4 – *VEGFA* copy number in different cancer tissues.

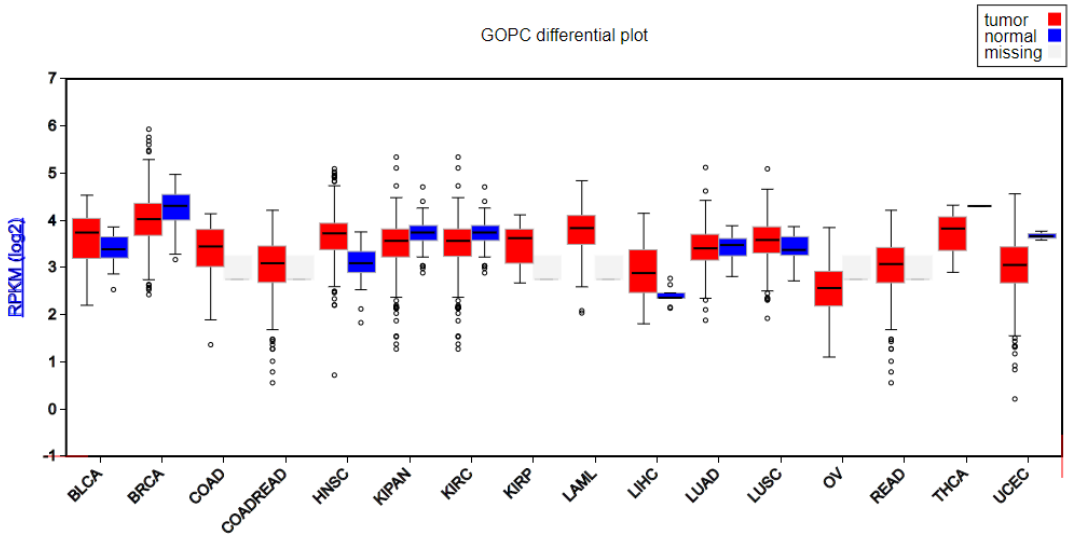


Figure 4.19: Target gene for MiR5 – *GOPC* copy number in different cancer tissues.

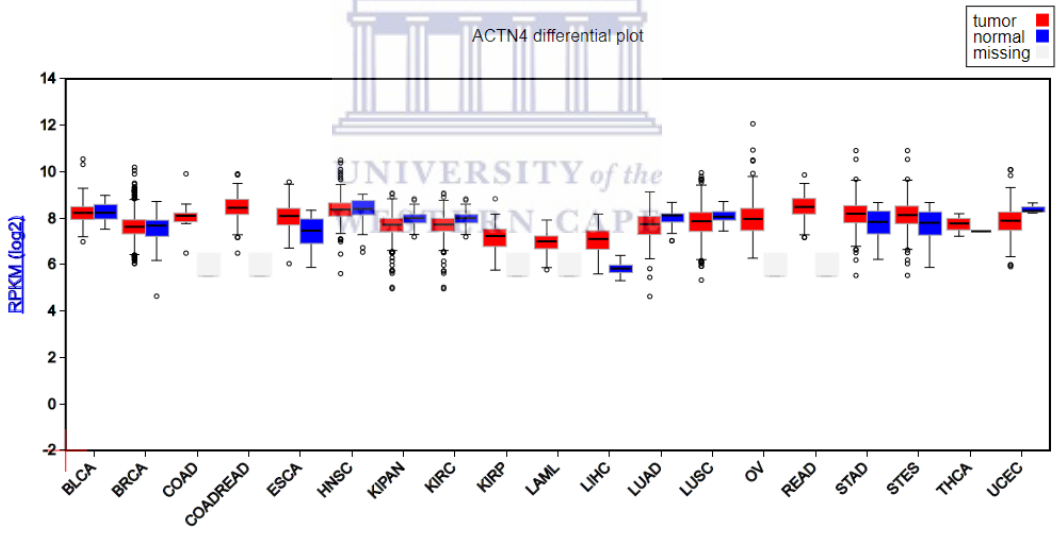


Figure 4.20: Target gene for MiR5 – *ACTN4* copy number in different cancer tissues.

4.6 Discussion

The results presented here demonstrate the expression of novel microRNAs as ways to identify their target genes in BC using *in silico* tools. Of the five microRNAs, four were expressed in BC tissue.

Maximum expression of MiR1 was obtained in breast tissue using miRmine, while its expression was also detected in several other organs. Furthermore, the expression of miR1 was up-regulated in serum samples of the BC database when compared with normal controls. However, it was absent in tissue samples. Based on this result, it is hypothesized that MiR1 could possibly be present as "circulating miRNA" and, therefore, possess great potential to serve as a diagnostic or prognostic biomarker. Solid tumors are known to secrete microRNAs which remain stabilized in body fluids, viz. blood, plasma and serum (Weber *et al*., 2010). In previous studies, several circulating microRNAs like miR-155, miR-195 and many others have been implicated as diagnostic and prognostic signatures in BC (Roth *et al.,* 2010; Heneghan *et al*., 2010; Wu *et al*., 2011; Hu *et al*., 2012). The up-regulated expression of MiR1 was also reflected by the over-expression of its target gene *HIP1* in almost all breast adenocarcinomas using the Oncomine database. Multiple studies reported a high expression of the *HIP1* gene in breast, colon, prostate, and other types of cancers (Rao *et al*., 2002; Rao *et al*., 2003). A study conducted by Rao et al.(2003) reported *HIP1* as a novel oncoprotein that transforms cells through dysregulation of multiple receptors involved in clathrin trafficking. Up-regulated *HIP-1* expression promotes

tumor formation and is associated with a general alteration in receptor trafficking (Rao *et al*., 2003). These results are in corroboration with previous results where the *HIP1* gene was highly expressed in low-risk BC patient data and is significantly correlated with prognosis and differential survival of BC patients (Chapter 3, sections 3.6.2 - 3.6.3). Together, MiR1 and its target gene showed elevated expression in BC and, therefore, can serve as important diagnostic and prognostic biomarkers.

MiR2 expression, as MiR1, was also found in all tissues including breast tissue. In the case of breast tissue, MiR2 expression was found to be considerably over-expressed in cancer serum samples when compared with normal controls. On the contrary, it was found in normal tissues alone but missing in cancer tissues. These results clearly indicate that as tumor progression occurs, expression of MiR2 might have increased with consequent up-regulated levels in the surrounding environment as a circulating microRNA. While MiR2 expression was remarkably increased in BC, similar outcomes were not reflected in its target gene, *SYT9*. From the present results, *SYT9* was revealed to be differentially expressed in multiple breast adenocarcinomas, with down-regulation of a few types and up-regulation of other types. This is again in agreement with our previous results where no prognostic significance of this gene was obtained in SurvExpress analysis, while an association with BC prognosis and survival outcomes were obtained in the Kaplan-Meier plotter outcomes (Chapter 3). Hence, it is suggested that although MiR2 has a good potential to serve as circulating

microRNA signature in BC prognosis and diagnosis, its expression needs requires validation.

Interestingly, although the genes (*TEMD10* and *AP1G1*) regulated by MiR3 showed differential expression in multiple breast adenocarcinomas, MiR3 on its own was absent in breast tissue samples. This is perhaps because of biases within microRNA-based human tissue-specificity created by differences in experimental conditions, nature of study and disease status of all tissues. Further *in-vitro* studies are needed to validate the microRNA expression in BC.

Similar to MiR1, the expression of MiR4 was significantly higher in breast tissue compared to other tissues. MiR4 expression was remarkably down-regulated in the serum of cancer patients alone while no noticeable change was observed in tissue samples. As far as expression of target genes of MiR4 is concerned, two genes *CD2AP* and *VEGFA*, were differentially expressed, while *ANP32E* was up-regulated in all types of breast adenocarcinoma. In our previous results, mixed outcomes regarding the significance of these genes as prognostic markers were also obtained and further studies are needed to confirm their utility as BC biomarkers.

The microRNA, MiR5, as MiR1 and MiR4, was also found to be expressed maximally in breast tissue in comparison with other tissues. When studied for its copy number expression in BC, its expression was marginally up-regulated in BC

tissue in comparison to normal controls, while it was absent in serum samples indicating its utility as a solid tumor marker. Additionally, the target genes of MiR5, *GOPC* and *ACTN4*, demonstrated reduced and differential expression in all types of breast adenocarcinoma. These results were in collaboration with previous outcomes where SurvExpress and Kaplan-Meier plotter outcomes identified *ACTN4* and *GOPC* individually as significant prognostic markers.

Considering the importance of microRNAs in the early diagnosis of BC for better disease prognosis, it is important to discover novel microRNAs and their target genes with diagnostic implications. The present study outcomes indicated that out of five novel microRNAs, MiR1 and MiR2 possessed the highest potential to serve as new prognostic and diagnostic signatures as circulating microRNAs. These microRNAs can be utilized as non-invasive biomarkers which can be used for easy disease diagnosis and prognosis. Similarly, MiR5 can act as a novel tumor marker whereas the remaining microRNAs need further validation.

4.7 References

Abdellatif, M., 2012. Differential expression of microRNAs in different disease states. *Circulation research*, *110*(4), pp.638-650.

Berindan-Neagoe, I., Monroig, P.D.C., Pasculli, B. and Calin, G.A., 2014. MicroRNAome genome: a treasure for cancer diagnosis and therapy. CA: a cancer journal for clinicians, 64(5), pp.311-336.

Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J.C., Hernandez-Boussard, T., Rees, C.A., Cherry, J.M., Botstein, D., Brown, P.O. and Alizadeh, A.A., 2003. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic acids research*, *31*(1), pp.219-223.

Dressman, M.A., Baras, A., Malinowski, R., Alvis, L.B., Kwon, I., Walz, T.M. and Polymeropoulos, M.H., 2003. Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. *Cancer Research*, *63*(9), pp.2194-2199.

Edelman, L.B., Eddy, J.A. and Price, N.D., 2010. In silico models of cancer. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(4), pp.438-459.

FIREBROWSE. 2016 Tutorial. Available at: http://firebrowse.org/tutorial/FireBrowse-Tutorial.pdf. 30/8/2017.

Glanzer, J.G. and Eberwine, J.H., 2004. Expression profiling of small cellular samples in cancer: less is more. *British journal of cancer*, *90*(6), p.1111.

Heneghan, H.M., Miller, N., Lowery, A.J., Sweeney, K.J., Newell, J. and Kerin, M.J., 2010. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Annals of Surgery*, 251(3), pp.499-505.

Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y. and Jian, T.Y., 2014. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic acids research*, 42(D1), pp.D78-D85.

Hu, Z., Dong, J., Wang, L.E., Ma, H., Liu, J., Zhao, Y., Tang, J., Chen, X., Dai, J., Wei, Q. and Zhang, C., 2012. Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls. *Carcinogenesis*, *33*(4), pp.828-834.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A. and Leiserson, M.D., 2013. Mutational landscape and significance across 12 major cancer types. *Nature*, *502*(7471), p.333.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp.27-30. Navarro, J.D., Niranjan, V., Peri, S., Jonnalagadda, C.K. and Pandey, A., 2003. From biological databases to platforms for biomedical discovery. *Trends in biotechnology*, 21(6), pp.263-268.

Lu, J., Getz, G., Miska, E.A. and Alvarez-Saavedra, E., 2005. MicroRNA expression profiles classify human cancers. *Nature*, *435*(7043), p.834.

Panwar, B., Omenn, G.S. and Guan, Y., 2017. miRmine: a database of human miRNA expression profiles. *Bioinformatics*, 33(10), pp.1554-1560.

Pruitt, K.D. and Maglott, D.R., 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic acids research*, 29(1), pp.137-140.

Rao, D.S., Bradley, S.V., Kumar, P.D., Hyun, T.S., Saint-Dic, D., Oravecz-Wilson, K., Kleer, C.G. and Ross, T.S., 2003. Altered receptor trafficking in Huntingtin Interacting Protein 1-transformed cells. *Cancer cell*, *3*(5), pp.471-482.

Rao, D.S., Hyun, T.S., Kumar, P.D., Mizukami, I.F., Rubin, M.A., Lucas, P.C., Sanda, M.G. and Ross, T.S., 2002. Huntingtin-interacting protein 1 is overexpressed in prostate and colon cancer and is critical for cellular survival. *The Journal of Clinical Investigation*, *110*(3), p351.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pander, A. and Chinnaiyan, A.M., 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*, 6(1), pp.1-6.

Roth, C., Rack, B., Müller, V., Janni, W., Pantel, K. and Schwarzenbach, H., 2010. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Research*, *12*(6), p.R90.

Sheedy, F.J., 2015. Turning 21: induction of miR-21 as a key switch in the inflammatory response. *Frontiers in Immunology*, *6*.

Subramanian, S., West, R.B., Corless, C.L., Ou, W., Rubin, B.P., Kent-Man, C., Leung, S.Y., Yuen, S.T., Zhu, S., Hernandez-Boussard, T. and Montgomery, K.,

2004. Gastrointestinal stromal tumors (GISTs) with KIT and PDGFRA mutations have distinct gene expression profiles. *Oncogene*, 23(47), p.7780.

Wang, X. and Wang, X., 2006. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Research*, 34(5), pp.1646-1652.

Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H., Lee, M.J., Galas, D.J. and Wang, K., 2010. The microRNA spectrum in 12 body fluids. *Clinical Chemistry*, 56(11), pp.1733-1741.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M. and Cancer Genome Atlas Research Network, 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, *45*(10), pp.1113-1120.

Wu, Q., Lu, Z., Li, H., Lu, J., Guo, L. and Ge, Q., 2011. Next-generation sequencing of microRNAs for breast cancer detection. *BioMed Research International*, *2011*.

Zhang, B., Pan, X., Wang, Q., Cobb, G.P. and Anderson, T.A., 2006. Computational identification of microRNAs and their targets. *Computational Biology and Chemistry*, *30*(6), pp.395-407.

CHAPTER 5

5. General conclusion and future work

5.1 Introduction

One of the leading causes of death worldwide, accounting for 8.8 million deaths in 2015 is cancer (WHO, 2017). Throughout the developed and developing world as including South Africa, cancer is a major killer (CANASA, 2017). There are multiple types of cancer including breast cancer (BC).

In 2012, Youl *et al*. (2016) reported that an estimated 1.67 million women had been diagnosed with BC worldwide. According to the 2011 National Cancer Registry (NCR), BC is not only the most common cancer in women of all races, but has a lifetime risk of 1:29 in South Africa (CANASA, 2017). Current diagnostic tools include BSE, mammography, ultrasound, and magnetic resonance imaging, which all have limitations like poor sensitivity and specificity, false-negative and false-positive results and are also invasive. Some of these techniques are also expensive. Therefore, new diagnostic biomarkers are under investigation, i.e. microRNA. The goal is always to diagnose cancers at the earliest possible stage providing patients with the best prognosis (Youl *et al.*, 2016).

These biomarkers for cancer must be acquired rapidly and reasonably with the use of non-invasive or minimally invasive procedures, amenable to a variety of

quantification techniques, and have a high sensitivity and specificity (Chatterjee *et al*., 2017).

Recently, microRNAs have emerged as one of the most important novel classes of cellular regulators, which control apoptosis, differentiation, and migration processes (Trang *et al*., 2017). They are omnipresent in at least 12 body fluids, i.e. serum, plasma, urine, and saliva, and as such provide an alternative for invasive biopsy procedures (Chatterjee *et al*., 2017).

There has been a growing interest among researchers to investigate the role of microRNA in normal cellular processes and disease. To assess the initiation and progression of tumours, and their response to treatment in cancer patients, expression profiling of microRNAs as diagnostic and prognostic biomarkers, has been introduced into cancer clinics. Researchers are focusing on circulatory microRNAs and evaluate their significance as diagnostic, prognostic and predictive biomarkers in cancer (Reddy & Kaladhar, 2015). MicroRNAs are differentially expressed depending on the molecular subtypes of tumours, whereby individual expression profiles classify tumour malignancies which play an important role in their development (Takahashi *et al*., 2015).

The usefulness of microRNAs has shown promise as a good biomarker for diagnostic, prognostic and possibly therapeutic purposes for many diseases as well as BC. Therefore, the aim of this study was to identity novel microRNAs as biomarkers

for early stage detection of BC. The study was divided into three parts, i.e. *in silico*, prognostic and predictive, and molecular identification. The *in silico* approach identifies novel microRNA within a short period of time, is rapid, cost-effective, and less labour intensive.

The objectives for the *in silico* study involved identifying microRNAs which were not previously found to be associated with BC. The *in silico* methodology makes use of several web-based tools for the prediction of microRNAs, their target genes and their mechanism of action.

5.2. *In silico* approach for identification of microRNAs as biomarkers for breast cancer detection (Chapter two)

This study successfully resulted in identifying nine target genes regulated by five microRNAs, which may play key roles in the BC development as well as progression. This study successfully identified five microRNAs (miR-1, miR-2, miR-3, miR-4 and miR-5) as well as their target genes, and thus may be used as biomarkers for the early detection of BC. MiR-4 and miR-5 and theirthree target genes, *VEGFA*, *CD2AP* and *ACTN4*, showed a clear link based on the STRING database, as described in section 2.12.4(Chapter) of this thesis. Additionally, as described in section 2.13, in the study by Honda, (2015) researchers found that *ACTN4* demonstrated not only a role in cancer invasion but also its biological role as a transcriptional co-activator. *ACTN4* is

aggressively involved in the tumorigenesis of BC, attracting a lot of attention. Furthermore, vascular *VEGFA* was reported to play a crucial role in tumour angiogenesis and cancer progression. *CD2AP* is fundamentally important in various cellular processes, which include endocytosis, apoptosis, cell survival, cell adhesion, structure and cytokinesis (Medway *et al*., 2013).

5.3 Prognostic and predictive validation of micoRNAs and their target genes as breast cancer biomarkers using an *in silico* approach (Chapter three)

In BC research, the discovery of prognostic and predictive biomarkers is a top priority undertaking. Therefore, it is of interest to validate prognostic or predictive candidate genes in appropriately powered BC cohorts (Györffy *et al*., 2010). The prognostic/predictive validation of these individual genes also showed variations as effective individual BC biomarkers.

The *HIP1* gene was be highly expressed in low-risk BC patient data, with no prognostic significance of this gene obtained in SurvExpress analysis. However, *AP1G1* was commonly identified as a significant prognostic biomarker in the initial analysis of two datasets. *TMED10* was highly expressed in the high-risk group. In SurvExpress analysis, *VEGFA* was identified as a significant prognostic marker with high expression of *CD2AP* in the high-risk group. *ACTN4* was also identified as a significant prognostic marker in the high-risk group.

In Kaplan-Meier plotter outcomes, the *HIP1* gene was found to correlate significantly with the prognosis and differential survival of BC patients. *AP1G1* prognostic importance was reproduced in Kaplan-Meier plotter analysis where the candidate biomarker gene showed good prediction of differential survival in BC patients. Kaplan-Meier plotter outcomes identified the prognostic importance of *GOPC* alone in the prediction of differential survival of BC patients.

In the *in silico* prognosis analysis of target genes of microRNA using GOBO analysis, all the nine target genes showed significant prognostic importance in predicting the survival outcomes in low-, intermediate- and high-risk patients. This was detected all oestrogen receptor positive tumour subtypes and oestrogen receptor negative luminal B tumour subtype.

5.4 Expression analysis of microRNA and target genes using *in silico* approaches(Chapter foure)

The database miRmine was used in the expression analysis of microRNA in tissue, and for its target genes, the Oncomine and FIREBROWSE databases were used.

Using miRmine, MiR1 expression–was found in breast tissue as well as in various other organs. Additionally, expression of the target gene, HIP1, for miR1 was found to be up-regulated in serum samples of BC from the database.

In all tissues as well as breast tissue, MiR2 expression was found to be over-expressed in cancer serum samples in comparison with normal controls. Expression of MiR2 was remarkably increased in BC. However, similar outcomes were not reflected in its target gene, *SYT9*.

MiR3 showed differential expression in multiple breast adenocarcinomas, but on its own was not detected in breast tissue samples, and thus not appropriate as a biomarker. However, to validate the microRNA expression in BC more *in-vitro* studies are needed.

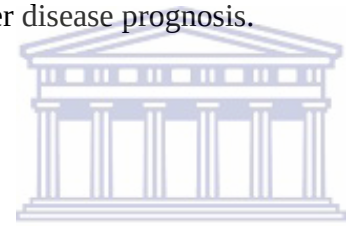
Compared with other tissues, MiR4 expression was more expressed in breast tissue, and remarkably down-regulated in the serum of cancer patients. The target genes, *CD2AP* and *VEGFA*, were expressed differentially, while *ANP32E* was up-regulated in all breast adenocarcinoma types.

In comparison with other tissues, MiR5 expression was the highest in breast tissue, whereas in BC tissue, its expression was marginally up-regulated and thus a sound tumor marker. MiR5 target genes, *GOPC* and *ACTN4*, revealed reduced and differential expression in all breast adenocarcinoma types.

5.5 Conclusion

Of the five novel microRNAs studied, MiR1 and MiR2 have the highest as new prognostic and diagnostic signatures. As non-invasive biomarkers they can be utilized for easy disease diagnosis and prognosis. MiR5 could also serve as a novel tumor biomarker, while further validation is required for the other two microRNAs studied. The significance of this study is that we did not find two microRNA target genes that could possibly be linked to BC.

In conclusion, early diagnosis of BC is important to discover novel microRNAs and their target genes for better disease prognosis.



5.6 Future work

Future work is suggested to investigate the evaluated expression profiles of microRNAs for the tissue using qRT – PCR. Additional analysis on a larger set of tissues to confirm their specificity also requires further investigation. the genes that are regulated by microRNA using luciferase assay and microRNAs that will be tested on BC tissue samples also need to be validated. Other molecular diagnostic techniques such as, FISH and IHC, which are invasive methods could be used to determine protein expression instead gene expression. These methods require a biopsy sample from the patient.

5.7 References

CANASA (2017) Prevalence Cancer, South African Cancer, Statistics "Accessed at http://www.cansa.org.za/south-african-cancer-statistics/ on 04/04/2017.

CANASA (2017) Women & Cancer "Accessed at http://www.cansa.org.za/womens-health/ on 04/04/2017.

Chatterjee, N., Rana, S., Espinosa-Diez, C. and Anand, S., 2017. MicroRNAs in Cancer: Challenges and Opportunities in Early Detection, Disease Monitoring, and Therapeutic Agents. *Current Pathobiology Reports*, *5*(1), pp.35-42.

Gmeiner, W.H. and Ghosh, S., 2014. Nanotechnology for cancer treatment. *Nanotechnology Reviews*, *3*(2), pp.111-122.

Győrffy, 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), pp.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), p.e82241.

Medway, C. and Morgan, K., 2013. CD2-Associated Protein (CD2AP). *In Genetic Variants in Alzheimer's Disease* (pp. 201-208). Springer New York.

Reddy, K.B., 2015. MicroRNA (miRNA) in cancer. *Cancer Cell International*, *15*(1), p.38.

Takahashi, R.U., Miyazaki, H. and Ochiya, T., 2015. The roles of microRNAs in breast cancer. *Cancers*, 7(2), pp.598-616.

Trang, P., Weidhaas, J.B. and Slack, F.J., 2017. MicroRNAs and Cancer. In *The Molecular Basis of Human Cancer* (pp. 277-286). Springer New York.

WHO (2017) Cancer Media Centre Factsheets, "Accessed at http://www.who.int/mediacentre/factsheets/fs297/en/ on 04/04/2017.

Youl, P.H., Aitken, J.F., Turrell, G., Chambers, S.K., Dunn, J., Pyke, C. and Baade, P.D., 2016. The Impact of Rurality and Disadvantage on the Diagnostic Interval for Breast Cancer in a Large Population-Based Study of 3202 Women in Queensland, Australia. *International Journal of Environmental Research and Public Health*, *13*(11), p.1156.