UNIVERSITY of the WESTERN CAPE

Bioinformatics Research Group

Characterization of ATP-binding cassette drug transporters and their role in breast cancer treatment using *in silico* approach

Mohammed Hashim Abdalraheem Hassan

A thesis submitted in fulfilment of the requirements for the degree of Magister of Science in the Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape

Supervisor: Dr. Ashwil Klein

Co-supervisor: Dr. Ashley Pretorius

December 2019

ABSTRACT

Breast cancer 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 breast cancer in their lifetime. In South Africa, breast cancer in women affects about 16.6 % of the population and could see a 78 % increase in cases by 2030. Comprehensive therapy on breast cancer including surgical operation, chemotherapy, radiotherapy, endocrinotherapy, etc. could help, but still has serious side effects. The Chemotherapy resistance against anticancer drugs is an emerging concern. Biomarkers have been identified as a viable option for early detection and progression of disease.

Examples of biological indicators for disease could be the ATP-binding cassette (ABC) drug transporters that utilizes the energy derived from ATP hydrolysis to efflux many chemically diverse compounds across the plasma membrane, thereby playing a critical and important physiological role in protecting cells from xenobiotics. These transporters are also implicated in the development of multidrug resistance (MDR) in cancer cells that have been treated with chemotherapeutics.

High expression of these membrane proteins as a family of ABC drug transporters are one of the main reasons for drug resistance by increasing the efflux rate of the anti-cancer drug from cancer cells. ABC drug transporters are considered to be one of the largest protein families in living organisms. There are 48 genes in the human genome that encode ABC transporters, which are divided into seven subfamilies (ABCA-ABCG). Studies revealed that ABC transporter genes has been shown to be associated with tumour development, progression and response to therapy, suggesting their possible use as diagnostic, prognostic and predictive biomarkers.

The aim of this study was to investigate and identify novel ABC transporter genes that could be implicated in breast cancer and MDR and potentially would be a therapeutic target for successful chemotherapy treatment and disease progression and survival in breast cancer patients.

An *in silico* approach was used to identify 10 ABC transporter genes (ABCB2, ABCB9, ABCB10, ABCC1, ABCC4, ABCC5, ABCC10, ABCC11, ABCC12, ABCD1) implicated in breast cancer by conferring drug resistance through over-expression in cancer cells. The *in silico* study investigated the tissue expression specificity, protein interaction/s, pathways, and comparative toxicogenomics of the identified ABC transporter genes using several computational software such as Tissue-specific Gene Expression and Regulation (TiGER), the Human Protein Atlas (HPA), Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), and The Comparative Toxicogenomics Database (CTD).

The 48 ABC transporter genes were shortlisted through very selective criteria that narrowed the genes down to 10. Differential expression analysis of the genes using TiGER and HPA compared expression in normal versus cancerous tissue of the candidate genes. The result showed that ABCC11 was preferentially expressed in breast tissue with an enrichment value higher than 10.0. The results also showed ABCC10 overexpressed in breast cancer tissue, making these two genes top candidates for further analysis.

Result from STRING database showed a strong functional interaction network between the prioritized genes through protein homology, co-expression and text mining as evidence for the observed interactions. Furthermore, the prioritized list of genes was submitted to the CTD for intersectional analysis to obtain the toxicity relationship between the genes and the Tamoxifen as the first line chemotherapeutic treatment for breast cancer. Venn diagrams obtained from CTD showed intersectional relation between ABCB2, ABCC1, ABCC4, ABCC11, and ABCD1 genes and Tamoxifen.

Furthermore, an *in silico* validation of the prognostic/predictive values of the 10 prioritized genes (list 2) was carried out using an online biomarker validation tool and database for cancer gene expression data using survival analysis (SurvExpress) and gene expression based survival analysis web application for multiple cancer (PROGGENE). Results obtained from the PROGGENE survival and predictive analysis showed good prognostic values for the genes ABCB2, ABCC1, ABCC4, ABCC10 and ABCC12 with their significance measured by the probability value (*Pv*) (0.053, 0.001118, 0.01286, 0.00604, 0.00157 respectively).

From this study ABCC1, ABCC4, ABCC5, ABCC10, and ABCC11 genes could serve as putative therapeutic target biomarkers for breast cancer treatment following further in depth analysis. However, the variance in the effectiveness of individual genes suggests that the set of genes would perform better than individual gene in the management of breast cancer. The modulating roles of ABCC4, ABCC5 ABCC10, and ABCC11 in drug induced apoptosis, suggest they could probably play an important role in personalized medicine and could serve as biomarkers to monitor the prognosis and/or therapeutic outcome of chemotherapy drugs in breast cancer patients.

The use of modern genomics, proteomics, bioinformatics, and systems biology approaches has resulted in a substantial increase in our ability to identify molecular mechanisms that are involved in MDR in cancer and to find drugs that may block or reverse the development of drug resistance. By using an *in silico* approach in this study, a list of five ABC transporter genes were identified, of which two (ABCC10 and ABCC11) could potentially serve as prognostic and predictive biomarkers for the management of breast cancer treatment.

KEYWORDS

ATP binding cassette drug transporter

Bioinformatics

Biomarker

Breast cancer

Chemotherapy resistance

Gene expression profiling

In silico

Prognostic analysis

Survival analysis

Therapeutic target validation

PLAGIARISM DECLARATION

I declare that *Characterization of ATP-binding cassette (ABC) drug transporters and their role in breast cancer treatment using in silico approach* 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.

Full name: Mohammed Hashim Abdalraheem Hassan.

Date: December 10, 2019

Signature: 

DEDICATION

For AFRICA, for SUDAN, for the new young generation who took it to the street bravely and throwing the most brutal and dictator muslim's brotherhood regime.

It was in the heart of the revolution when I was writing this thesis, I could not even sometimes stay focus to do my work. But the people well has overcome the oppressive regime.

ACKNOWLEDGEMENT

I would like to thank the **National Research Foundation** (NRF) for their valued support in the form of the Innovation Scholarship that was given to me to further pursue my master degree at the University of the Western Cape, South Africa. I am truly thankful and grateful for the great opportunity and it would not have been possible for me to do my research without this scholarship.

I would like to express my sincerest gratitude to my supervisors, **Dr. Ashwil Klein** and **Dr. Ashley Pretorius**, for their 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 both of you for proofreading my thesis. I could not have asked for better supervisors.

For my support system, **Miss. Mischka Lewis**, the one who has always been next to me since I started the journey of my two year's master research, creating a conducive environment for me to study and handling the stressful time, especially during writing my thesis. I would never find a word or a sentence to be used to appreciate your huge support in accomplishing this study on time. I thank you for all the support and help and I would forever be appreciative for the chance that we have met in this life and we are together and holding our dreams together. I love you habibty.

I would go as far as to the New Mexico, United State of America, **Dr. Sultan Hassan.** My beloved brother, who has been an immense inspiration and support towards achieving my research. I am thankful to you for believing in my abilities and capabilities in achieving this studies and allowing me to take my chance to explore my own way. Thank you for encouraging me to pursue my postgraduate studies and build my academic career. I wish you all of the best for you and your dearly loved family and I cannot wait to see the new member to the family "Neel".

For **Dr. Nicole**, I stay appreciative and humble to thank you. You have your doors always open for assistance, guidance and help. I am forever grateful and thankful for all the support and help that has contributed positively in my research journey. I thank you Sister Nicole.

For **Dr. Chipampe**, it's been a great time working with and under your kind guidance throughout my research and you have been a greater support and guidance. I am appreciative for the valuable discussion and input into my research and it has been the most incredible. I am truly thankful Dr. Chipampe.

For **Dr. Abdulrahaman Elbagory**, Abdul for short, you have been a big brother and good friend for me during this hectic journey, I remember all those nights that we spent in the lab even during your Ph.D. journey which has come to an end. I would like to thank you for all of the support and help in finalizing my thesis and making sure that everything is fine in my thesis. I am genuinely appreciative for the time that we firstly met at Hector Petersen Residence in 2016 and from that time you have shown loyalty to our friendship. I appreciate you in my life. I wish you all of the best with your academic endeavors and you are a true Nanotechnologist and am quite sure that you will definitely make significant changes in nanotechnology world.

To **DR. Martin**, these are your initials (DR) but soon you will be conferred your Ph.D. degree and I cannot wait to celebrate it out loud DR. Riziki Martin. You are an amazing and genuine person that I have met in this life. Polite, honest and most importantly very helpful. I thank you for all of the support and help during my master degree and you have been always a good friend.

To the members of the Biotechnology department, I can't name you all, but I thank you for everything you have shared with me these years. Thank you to the University of the Western Cape for allowing me to complete this part of my education at the institution.

I would also like to say a heartfelt thank you to my mother "**Karima**" and father "**Hashim**" for always believing in me and encouraging me to follow my dreams. And my sisters: **Dr. Zainab**, **Dr. Hanaa**, **Mrs. Salma**, **Miss. Fatima**, and the beloved Sister **Abeer**, and my brothers: **Eng. Abdalraheem** and **Dr. Sultan** for helping in whatever way they could during this challenging period.

I praise God, 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.

LIST OF ABBREVIATION

5dUMP	5-fluoro-2'-deoxyuridine 5'-monophospate
5-FU	5 Fluorouracil
ABC	ATP-Binding Cassette
ACS	American Cancer Society
ALDP	Adrenoleukodystrophy
BCRP	Breast Cancer Resistance Proteins
BRCA-1	Breast Cancer type 1
BRCA-2	Breast Cancer type 2
CA 15-3	Cancer Antigen 15-3
CAGE	Cape Analysis of Gene Expression
CANSA	Cancer Association of South Africa
cDNA	Complementary Deoxyribonucleic Acid
CIS	Carcinoma *in situ*
CRM	Cis-regulatory Module
CTC	Circulating Tumor Cells
CTD	Comparative Toxicogenomics Database
DCIS	Ductal Carcinoma *in situ*
DNA	Deoxyribonucleic Acid
DSBR	Double-strand Break Repair
EGFR	Epidermal Growth Factor Receptor

ER	Estrogen Receptor
EST	Expressed Sequence Tags
FISH	Fluorescent *in situ* Hybridization
GOBO	Gene Expression-Based Outcome for Breast Cancer Online
HER-2	Human Epidermal Growth Factor Receptor 2
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome
HPA	Human Protein Atlas
IDC	Invasive Ductal Carcinoma
IDs	Identities
IHC	Immunohistochemistry Assay
ILC	Invasive/ Infiltratory Lobular Carcinoma
Ki67	Antigen KI-67
MDR	Multidrug Resistance
MDR1	Multidrug Resistance 1 Protein
miRNA	MicroRNA
mRNA	Messenger Ribonucleic Acid
MRP1	Multidrug Resistance Protein1
NAC	Neoadjuvant Chemotherapy
NBDs	Nucleotide Binding Domains
NCBI	National Center for Biotechnologies Information

NCR	National Cancer Registry
NSCLC	Non-small Cell Lung Cancer
PR	Progesterone Receptor
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RNA	Ribonucleic acid
SLC	Solute Carrier
STRING	Search Tool for the Retrieval of Interacting Genes
SurvExpress	Online biomarker validation tool and database for cancer gene expression data using survival analysis
TiGER	Tissue-specific Gene Expression and Regulation
TM	Transmembrane
TMDs	Transmembrane Domains
WHO	World Health Organization

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

LIST OF TABLES

CHAPTER ONE

CHAPTER TWO

TABLE OF CONTENT

CHAPTER ONE: Rationale and Literature Review

1.1. Introduction on cancer

Cancer has become the leading cause of death worldwide, according to the World Health Organization (WHO), it is projected that death resulting from cancer will rise above 13.1 million by 2030 (Mathers and Bonita, 2009). The American Cancer Society (ACS) in 2019, estimated that 268,600 new breast cancer cases were diagnosed and 41,760 cases of breast cancer related deaths occurred (ACS, 2019). According to the registry of the Cancer Association of South Africa (CANSA), there is one in six male and one in nine female cancer patients (Ahmedin *et al*., 2011).

The understanding of cancer formation, development and its consequences as a disease has changed over the past decades, due to the improvement in treatment and the technological advancements that has been applied. Cancer is a cluster of diseases which could be described by its abnormality and uncontrolled growth of cells with the potentiality to spread or invade to other parts of the body (Hahn and Weinberg, 2014). Genetically, cancer arises as a result of a loss of several pathways in normal cells that control their growth and undergoes a series of molecular events that eventually change the normal growth pathways of a cell. A common phenotypic abnormality of cancer cells is the impairment of regulation of cell cycle control (Grumolato and Aaronson, 2014). The metamorphosis of a normal cell to a cancer cell is speculated to be in relation to many factors. However, the key factor is the mutational changes in genes that normally contribute to controlling of cell cycle progression which ultimately will lead to loss of the cell growth regulatory machinery (Zhou *et al*., 2013).

The affected genes are divided into two broad categories. Oncogenes are genes that promote cell growth, division, and reproduction. Whereas, tumour suppressor genes regulate a cell during the cell division and replication. This increases cell survival by controlling the cell growth, and if the tumor suppressor gene is not functioning properly, the cell can grow out of control, which could lead to a cancer (Grumolato and Aaronson, 2014). Cancer develops due to extensive genetic abnormalities and deviation in gene expression patterns. As it is described below in figure 1.1, in normal tissues, cells are largely instructed to grow by their neighbors or *via* systemic signals and the rate of new cell growth and old cell death are kept in balance. In contrast, in cancer tissues, there is a loss of balance between cell division and cell death. This disruption can result from uncontrolled cell growth or loss of a cell's ability to undergo apoptosis (Allinen *et al*., 2004).

Apoptosis, or programmed cell death, is the mechanism by which old or damaged cells normally self-destruct (Karvar, 2014).

Tumours are classified into two forms, which can either be malignant or benign. A benign tumor is usually removable and do not possess the ability to spread to other parts of the body which can also be described as a non-cancerous tumor (Hahn and Weinberg, 2014) A malignant tumour can be described as a metastatic tumour that has the ability to invade and cause damage to nearby tissues (Fletcher *et al*., 2010). Due to their malignant nature, the cells metastasize by breaking away and entering the bloodstream and lymphatic system, causing widespread destruction by forming tumours elsewhere in the body (Al-Mansouri and Alokail, 2006).

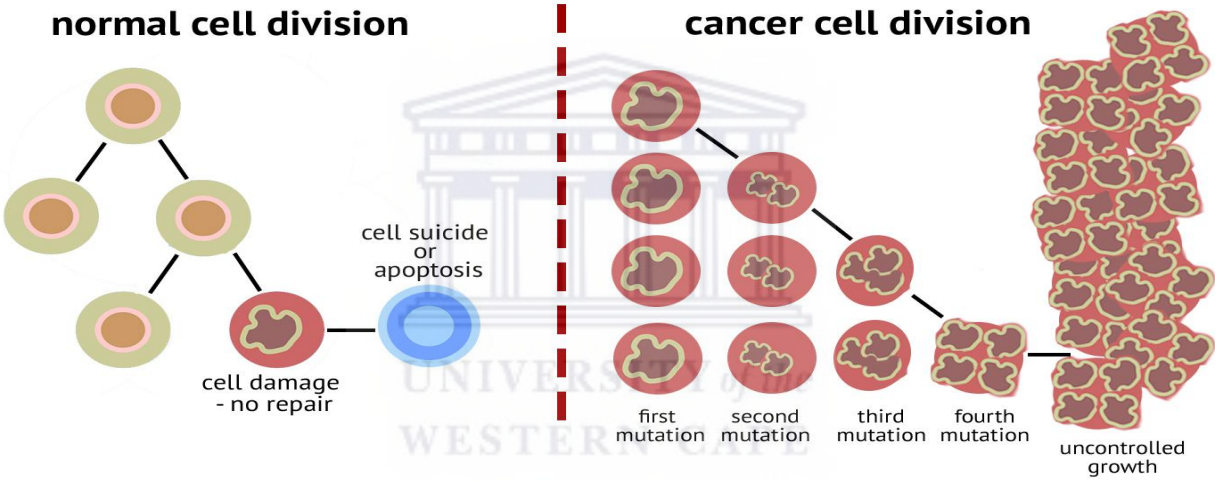

Figure 1.1: Represents the typical normal and cancer cell division and their differences.

1.1.1. Hallmarks of cancer

Cancer growth emerges in practically all human tissues and it is acknowledged that the fundamental processes that change a normal cell into a cancer cell are basically similar in all tumours developing in the human body irrespective of the initial site of development. These essential properties of survival, proliferation and dissemination are known as the hallmarks of cancer (Hanahan and Weinberg, 2000). In spite of the fact that these hallmarks are believed to be normal for a wide range of diseases, they are obtained through various distinct systems during numerous occasions of the multistep tumourigenesis process in various types of malignant growth.

More than 10 years ago, Hanahan and Weinberg highlighted six hallmarks of cancer, which are functional abilities that empowers a tumor to multiply, endure, spread. These hallmarks include: I) continuing proliferative signaling, ii) evading development suppressors, iii) activating invasion and metastasis, iv) enabling replicative immortality, v) angiogenesis and vi) resisting cell death (Hanahan and Weinberg, 2011). Globally, it's recognized in a few and maybe all kind of cancers (see figure 1.2) (Cook *et al*., 2005; Pietras and Östman, 2010).

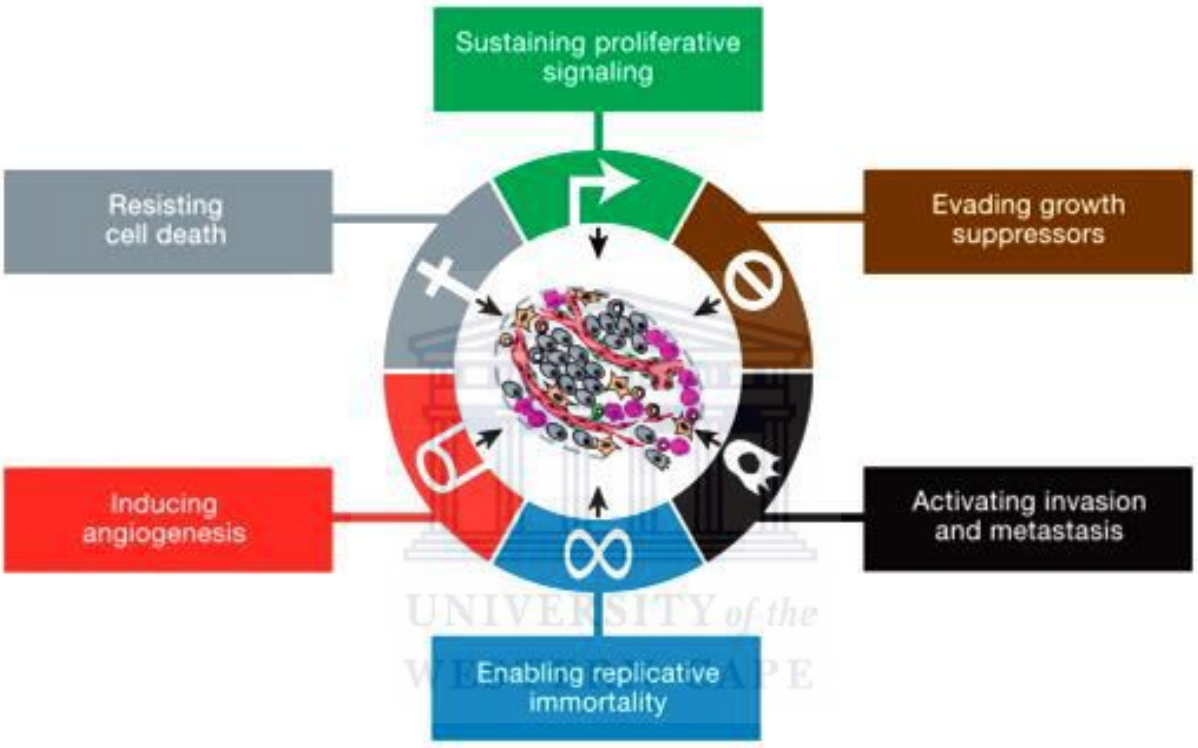

Figure 1.2: The six hallmarks of cancer acquired by normal cells as they evolve to a neoplastic state (Adapted from Hanahan and Weinberg, 2011).

The high throughput advancement of research increased the understanding of these hallmark capabilities with new pathways discovered as the molecular basis of cancer progression and treatment response and adding four more characteristics, which are described below in figure 1.3: (1) Deregulating cellular energetics; (2) avoiding immune destruction; (3) tumour promoting inflammation and (4) genome instability and mutation (Gutschner and Diederichs, 2012).

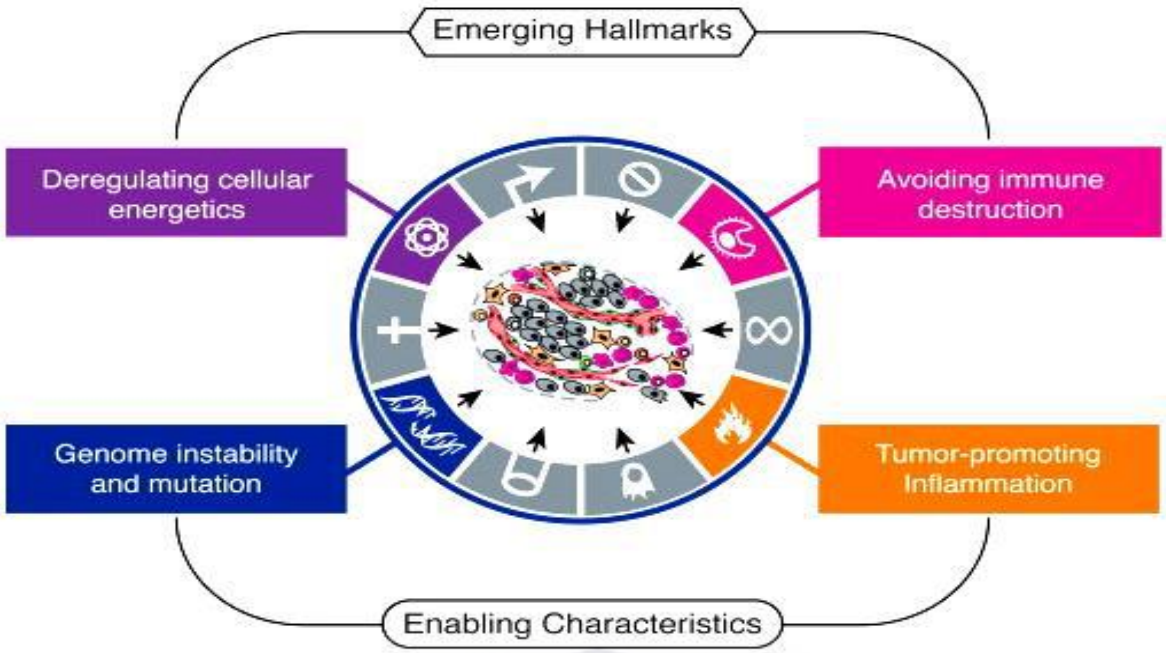

Figure 1.3: Emerging Hallmarks of cancer and Enabling Characteristics (Adapted from Hanahan and Weinberg, 2011).

1.2. Cancer risk factors

In developing countries, cancer has become a known public health challenge. This is partly associated with the aging and growth of the population and increased popularized cancer risk factors (Oeffinger *et al*., 2015). In Africa, the increased prevalence of cancer risk factors is associated with economic transitions as well as the adoption of western lifestyles. According to the United Nation's Population estimations, a 50 % and 90 % increase in the emergence of new cancer cases in the elderly in 2010 and 2030, respectively, is anticipated (Ahmedin *et al*., 2011). 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).

Factors that contribute to the development or progression of cancers are either environmental (external) or hereditary (internal) (Mathers and Bonita, 2009). Environmental factors include tobacco, chemicals, radiation, infectious organisms etc., whereas hereditary factors includes hormones, inherited mutations, immune conditions and mutations that develop from metabolism

(Ahmedin *et al*., 2011). The incorporation of both factors play a major role in the development of genetic malformation. The abnormalities in cancer cells usually result from mutations in protein encoding genes that regulate cell division, which can lead to the loss of cell growth regulation and by time, genes could become mutated (Craig Venter *et al*., 2001).

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 % of female deaths from lung cancer (Moodley *et al*., 2016). Obesity is another main risk factor for developing chronic diseases, including malignancies and several types of tumours such as breast, prostate, and kidney cancer (Montanari and Ecker, 2015, Slaoui *et al*., 2014).

1.3. Genetics of cancer

Cancer is a collection of diseases involving modification in the status and expression of multiple genes that confer a survival advantage and undiminished proliferative potential to somatic or germinal cells. The gene expression alterations are categorized into three main types of genes which are: (proto) oncogenes, tumour suppressor genes and deoxyribonucleic acid (DNA) repair genes. Together these main types of genes contribute to the development of the cancer genotype and phenotype that resists the natural and inherent death mechanisms established in cells (e.g. apoptosis), incorporated with dysregulation of cell proliferation events (Yang *et al*., 2015).

A normal cell needs to undergo several changes in order to become a cancer cell. Of these changes, genetic polymorphism variation as well as the gene expression regulation and alteration are considered to be key factors. As a result of these changes, cancer can therefore be classified as a genetic disease on the cellular and inheritance level (Hahn and Weinberg, 2014). It is estimated that about 35000 genes in the human genome have been known to be implicated to have some sort of association with various cancers (Yasui *et al*., 2004).

As mentioned above, mutations in three classes of genes can help to promote the formation of cancers in humans. However, mutations in genes result in altered proteins, such changes could occur during cell division or as a result of interaction with external agents or a random event of regulating a gene expression (Bertoncini *et al*., 2011). Some of these changes include physical loss of DNA, changes in nucleotides as well as epigenetic effects (Stenson *et al*., 2012). Most cancers result from either somatic cell mutations or germ-line cell mutations (Bertoncini *et al*., 2011).

However, Mutations in cancer cells, even if thought of as being insignificant, are candid and have been found to be present in all somatic cells of affected individuals (Balaji *et al*., 2016a). These mutations are not only known to generate cancer but are also seen as precursors or biological markers for disease as they can also be passed down to subsequent generations. In almost all cases approximately 90-95% of cancers are sporadic and the remaining 5-10% is inherited (Estudante *et al*., 2016; Stenson *et al*., 2012).

In mammary carcinomas, several factors are linked to its progression and appear to influence patient's survival rates. These biological molecular markers might claim reconsideration if individual genetic modification can be targeted by means of specific therapies (Ota *et al*., 2010). These markers have been classified into two groups depending on their role in the development and progression of cancer; oncogenes and tumor suppressor genes. An oncogene results from a gain-of-function mutation of a proto-oncogene that lead to a tumorigenic product, whereas mutation of a tumor suppressor causes a loss-of-function in the ability to control cell growth (Ota *et al*., 2010).

1.3.1. Oncogenes

Oncogenes encode proteins that possess the ability to cause cellular transformation. These genes act in a dominant way, either through overexpression or activating mutations (Grumolato and Aaronson, 2014). There are several criteria that define cellular transformation. These include morphologic changes, loss of contact inhibition, anchorage-independent growth, and the ability to form tumors. For example, under normal physiologic situations, a growth factor binding to a receptor produces a very transient activation of a certain signaling cascade allowing very tightly regulated responses such as proliferation to occur (Wang *et al*., 2011). When downstream components of these cascades are mutated in a way that causes them to be constitutively active, the signal is no longer transient and regulated, but is aberrantly turned on in a continuous fashion (Grumolato and Aaronson, 2014).

In addition to activating mutations, these genes can be activated by over expression at levels much higher than in normal cells. Proto oncogenes are commonly involved in cellular signaling (Tabarestani *et al*., 2014). The oncogenes suffer mutations that deregulate or activate their protein products so that they function at higher levels, activities, or at inappropriate times and places in a cell (Grumolato and Aaronson, 2014). Genetic variation in oncogenes act in a dominant fashion

meaning that, only one of the two alleles in a cell is commonly affected (Doyle and Ross, 2003). However, these mutations can either be gene amplifications or promoter mutations that increase the levels of a protein. Another type of these mutation would be the gene translocations, that produce fused and altered proteins or missense point mutations at selected places in a gene and its protein which would lead to activate an activity or produce a protein that cannot be properly regulated or degraded (Nalejska *et al*., 2014). Growth factor receptors, protein kinases, G-protein signaling molecules, and transcription factors in selected signal transduction pathways are common targets for oncogene mutations (Stenson *et al*., 2012).

1.3.2. Tumor suppressor genes

A tumor suppressor gene, or anti-oncogene, is a gene that protects a cell from one step on the path to cancer. When this gene mutates to cause a loss or reduction in its function, the cell can progress to cancer, usually in combination with other genetic changes (Levine *et al*., 2014). The loss of these genes may be even more important than proto oncogene or oncogene activation for the formation of many kinds of human cancer cells (Tian *et al*., 2016).

Tumor suppressor genes can be grouped into categories including caretaker genes, gatekeeper genes, and landscaper genes. The classification schemes are evolving as medicine advances, learning from fields including molecular biology, genetics, and epigenetics (Hahn and Weinberg, 2014). Tumor suppressor genes commonly contribute to the fidelity of the cell cycle replication process. They may act as negative regulators of oncogenes, cell cycle check points, or gene products that supply the appropriate nutrients or components to complete a faithful cell cycle division in the absence of stress (Levine *et al*., 2014).

Mutations in tumor suppressor genes are loss-of-function mutations and so occur in both alleles of a gene. Deletions, nonsense mutations, frame-shift mutations, insertions, or missense mutations that inactivate functional activity of a protein are all observed in tumor suppressor genes (Nalejska *et al*., 2014). Some tumor suppressor genes have a haplo insufficient phenotype which indicates that the reduced levels of proteins in cells that lack one allele of the genomic locus results in the inability of the cell to execute normal cellular functions contributing to tumor development.

In animals or humans with one mutant allele and one wild-type allele of a tumor suppressor gene, a suboptimal level of the gene product results in a lower level of function and a loss of fidelity

(Levine *et al*., 2014). For example, the p53 tumor suppressor gene in the heterozygous condition (Li-Fraumeni syndrome) has a lower level of apoptosis in lymphocytes exposed to stress in both mice and humans (Li-Fraumeni syndrome) when compared to two wild-type copies of that gene (Allinen *et al*., 2004). It is rare that the same gene in the same cell of an organism is mutated two independent times. Rather tumor suppressor genes accumulate two mutations in the same gene by a process termed "reduction to homozygosity" which is mediated by either gene conversion (*via* replication or recombination) or loss of one chromosome (with the wild-type allele) and duplication of the chromosome with a mutant allele on it. The net result of these events are two mutant alleles of a tumor suppressor gene (Levine *et al*., 2014).

1.4. Breast cancer

Breast cancer refers to a metastatic tumour that has developed from cells in the breast. Usually the malignancy either begins in the lobules cells which is the milk-producing glands, or the ducts, which are the passages that transfer milk from the lobules to the nipple for breast feeding (Fanale *et al*., 2012; Moodley *et al*., 2016; Russo *et al*., 2000). Breast cancer can also start in the stromal tissues, which include the fatty and fibrous connective tissues of the breast, although this is rare (Devadoss *et al*., 2013). Among solid cancers, breast cancer represents perhaps one of the best fields to study due to clinical drug resistance and of its reversal (Doyle and Ross, 2003).

Breast cancer start when cells in the breast begin to grow uncontrollably, these cells usually form a tumour that can often be seen on an x-ray or felt as a lump. The tumour is malignant (cancerous) when the cells grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body (Leonard, 2003; Devadoss *et al*., 2013).

The cancer cells can spread to nearby breast tissue with time and make their way into the small organs that filter out foreign substances in the body known as the lymph nodes. The lymph nodes then act as a pathway for these cancer cells into other parts of the body (Devadoss *et al*., 2013). The metastasized cancer cells may attach to other tissues and grow to form new tumours of those tissues. For breast cancer, the most common sites for metastatic tumours are the bones, liver, lungs, and the brain (Butti *et al*., 2019).

According to the WHO, breast cancer is the top cancer in women worldwide and is increasing particularly in developing countries where the majority of cases are diagnosed at the late stages

(WHO, 2019). Lack of early detection programmes and access to treatment place lead women in less developed regions to a high mortality risk.

Breast cancer is the most common, frequent and invasive cancer occurring in women across the world, with approximately 1.7 million new cases diagnosed in 2012, with a global estimation of 25% in new diagnosed cases of all cancer cases (Balaji *et al*., 2016a). According to the ACS in 2019, it is estimated that about 268,600 new cases of invasive breast cancer will be diagnosed in women and 41,760 death cancer cases will be from breast cancer (ACS, 2019).

Breast cancer is the most common form of cancer among women in South Africa, with an age-standardized incidence rate of 27 per 100 000 women, and a major cause of cancer mortality, accounting for 16% of cancer deaths among women in 2016 (Moodley *et al*., 2016). According to the National Cancer Registry (NCR) of South Africa, in 2010, breast cancer was reported in 3,157 cases amongst Black population, 816 cases amongst Colored population, 1,824 cases amongst White population and 340 cases amongst Asian population (Cancer Association of South Africa, 2015). In addition, breast cancer is the most common cancer in pregnant and postpartum women, occurring in about 1 in 3,000 pregnant women (Moodley *et al*., 2016).

In South Africa, geographically, the Western Cape and Free State had the highest rates of new cancer cases in 2017 (339 per 100000 for the Western Cape and 333 per 100000 for the Free State). Limpopo and the Eastern Cape had the lowest incidence rates of 217 and 292, respectively. According to statistics from the NCR 2014, the top five cancers affecting women in South Africa include: breast, cervical, colorectal, and uterine and lung cancer. Both breast and cervical cancer have been identified as a national priority with increasing incidences occurring.

For women, breast cancer is the most prevalent across the board, with Discovery insurance scheme reporting an incidence of 128 per 100000 in 2017. In second and third place for Discovery are colorectal cancer (27 per 100000) and lung cancer (21 per 100000). These top three are the same as the global top three. In South Africa, however, in second and third place to breast cancer, are cervical/uterine cancer and lung cancer.

In figure 1.4, approximately 19.4 million women aged 15 years and older live at-risk of being diagnosed with breast cancer. In 2013, deaths from breast cancer and cancers of the female genital tract, accounted for 0.7% and 1% of all deaths in South African, respectively. Statistically in 2018

(Figure 1.5), new cases of breast cancer were 2.500.000 million and resulting in 424,000 deaths in South Africa (WHO, 2018).

Figure 1.4: Represent age-standardized cancer mortality and cancer incident charts (NCR, 2017).

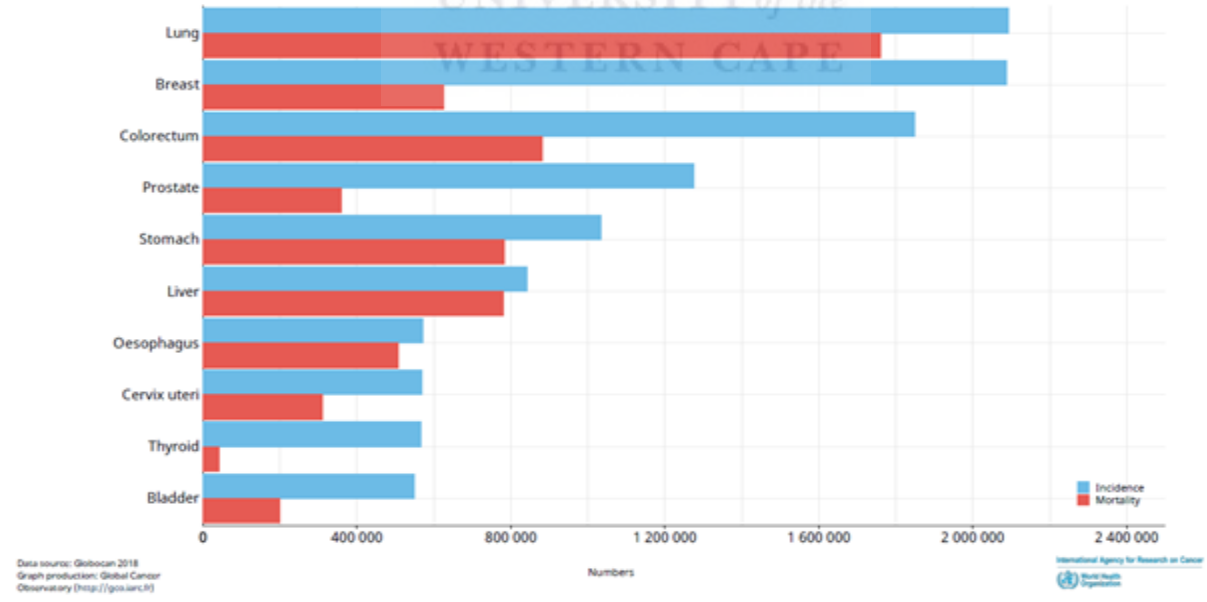

Figure 1.5: Represents estimated number of incident cases (in blue) and mortality cases (in red) worldwide, both sexes, all ages (Adopted from WHO 2018).

1.4.1. Anatomy and normal function of the female breast (Mammary Gland)

The mammary/breast gland is a unique organ to the class Mammalia, with the specific function to synthesize, secrete, and deliver milk to the newborn upon demand for its optimal nourishment, protection, and development (Hassiotou and Geddes, 2013). In humans, the life cycle of the female mammary gland is epitomized by drastic changes in composition, architecture, and functionality, mediated by marked changes in gene expression, that characterize its physiological stages of development, all of which are aimed at allowing it to perform its function as a milk-producing organ with the birth of the infant (Hassiotou and Geddes, 2013).

Given this function, it is only during a pregnancy/lactation cycle that the gland reaches a mature developmental state through hormonal influences at the cellular level that effect drastic modifications in the micro- and macro-anatomy of the gland, resulting in remodeling of the gland into a milk-secretory organ (Mills *et al*., 2011). Pubertal and post-pubertal development of the breast in female's aids in preparing it to assume a functional state during pregnancy and lactation. Human breast milk has a unique biochemical and cellular composition, providing the infant with optimal nutritional, protective, and developmental factors (Mills *et al*., 2011).

The breast is an organ that reflects its special function by its structure which is the production of milk for the newborn referred to as lactation (breast feeding). The epithelial component of the tissue consists of lobules, where milk is made, which connect to ducts that lead out to the nipple. A lobule is a gland that primarily functions to make milk in the breast. A collection of lobules along with a small duct make up the terminal ductal lobular units of the breast. Whereas the duct is a tubular structure that carries milk from the terminal ductal lobular unit to the nipple (Hassiotou and Geddes, 2013).

Most cancers of the breast arise from the cells which form the lobules and terminal ducts. These lobules and ducts are spread throughout the background fibrous tissue and adipose tissue (fat) that make up the majority of the breast. The male breast structure is nearly identical to the female breast, except that the male breast tissue lacks the specialized lobules, since there is no physiologic need for milk production by males (Hassiotou and Geddes, 2013).

Anatomically, the adult breast sits atop the pectoralis muscle (the "pec" chest muscle), which is atop the ribcage. The breast tissue extends horizontally (side-to-side) from the edge of the sternum

(the firm flat bone in the middle of the chest) out to the midaxillary line (the center of the axilla, or underarm). A tail of breast tissue called the "axillary tail of Spence" extends into the underarm area. This is important because breast cancer can develop in this axillary tail, even though it might not seem to be located within the actual breast (Litviakov *et al*., 2016).

The breast tissue is encircled by a thin layer of connective tissue called fascia. The deep layer of this fascia sits immediately atop the pectoralis muscle, and the superficial layer sits just under the skin. The skin covering the breast is similar to skin elsewhere on the body and has similar sweat glands, hair follicles, and other features. A clinician will examine the skin in addition to the breast tissue itself when performing a breast examination (Hassiotou and Geddes, 2013).

The blood supply from the breast comes primarily from the internal mammary artery, which runs underneath the main breast tissue. Blood vessels and capillaries are housed within the mammary stromal matrix delivering biochemical and cellular components essential for the function of the gland and milk synthesis (Waks and Winer, 2019). The blood supply provides much needed gases and nutrients, such as oxygen, iron, zinc and vitamins to the breast tissue. The lymphatic vessels of the breast flow in the opposite direction of the blood supply and drain into lymph nodes. It is through these lymphatic vessels that breast cancers metastasize or spread to lymph nodes. Most lymphatic vessels flow to the axillary (underarm) lymph nodes, while a smaller number of lymphatic vessels flow to internal mammary lymph nodes located deep in the breast (Hassiotou and Geddes, 2013).

Understanding of this lymphatic drainage is important, because when breast cancer metastasizes, it usually involves the first lymph node in the chain of lymph nodes. This is called the "sentinel lymph node," and a surgeon may remove this lymph node to check for metastases in a patient with breast cancer. Many additional changes are seen in the breast tissue during pregnancy and lactation due to the changes in hormones during those times (Hassiotou and Geddes, 2013; Waks and Winer, 2019).

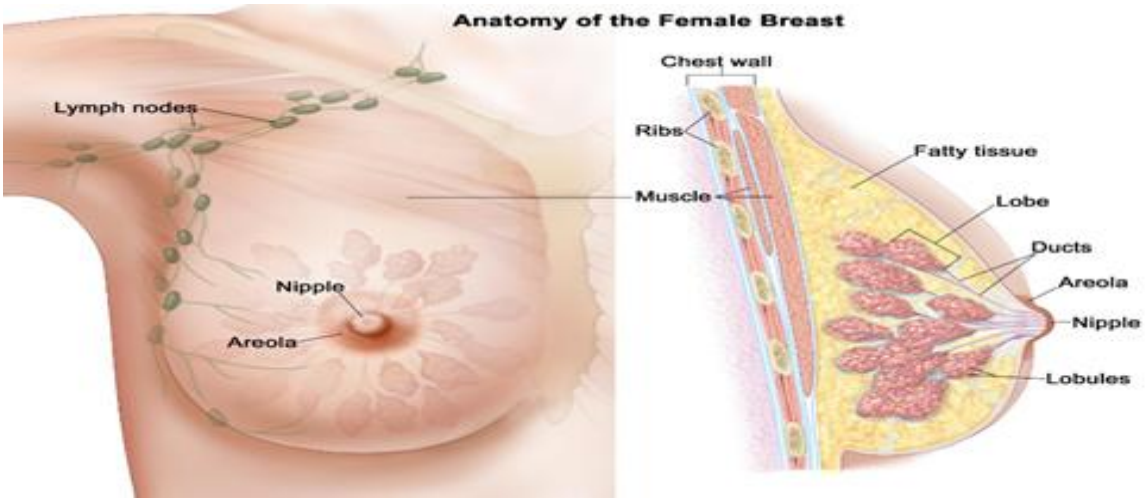

Figure 1.6: Anatomy of the female breast (adopted from Hassiotou and Geddes, 2013).

1.4.2. Types of breast cancer

Breast cancer can be categorized into *in situ* and invasive (infiltrating) carcinoma based on the histological and morphological appearance of cancer cells. Breast carcinoma *in situ* (CIS) can either be ductal or lobular (Allinen *et al*., 2004).

Ductal carcinoma *in situ* (DCIS) is 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, lymphatics, blood vessels, stromal cells, immune cells and fat cells can either promote or suppress the carcinogenic process by responding to secretory signals. 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 (Allinen *et al*., 2004; Bravaccini *et al*., 2013).

IDC originates from the milk ducts and spreads through the walls of the ducts to surrounding breast tissue. It is the most common type of breast cancer accounting for 72 to 80 % of all invasive breast cancer and 8 to 14 % of all breast cancer cases (Arps *et al*., 2013). Invasive carcinoma is divided into six types namely, tubular, ductal lobular, infiltrating ductal, mucinous, medullary and invasive or infiltratory lobular carcinoma (ILC). 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 (Allinen *et al*., 2004; Arps *et al*., 2013; Bravaccini *et al*., 2013).

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. 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 (Allinen *et al*., 2004; Arps *et al*., 2013; Moodley *et al*., 2016).

1.4.3. Risk factors of breast cancer

Major risk factors of breast cancer 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 breast cancer 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 breast cancer progression and are currently used as molecular markers for diagnostic and therapeutic intervention in breast cancer patients (Petrucelli *et al*., 2019).

Women who have started menstruating early in life or who have late menopause were reported to be at an increasing risk of developing breast cancer. An increased risk of breast cancer 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 (Kotsopoulos *et al*., 2017). Regarding the effect of pregnancy on breast cancer, neither diagnosis of breast cancer during pregnancy nor pregnancy after breast cancer seems to be associated with adverse survival outcomes in women who carry a BRCA1 or BRCA2 mutation (Kotsopoulos *et al*., 2017; Petrucelli *et al*., 2019).

Hormonal history for women appears to be a risk factor, as the relative risk of breast cancer seems to be related to the breast‘s cumulative exposure to estrogen and progesterone (Hamilton and Piccart, 2000; Hulka and Moorman, 2001). Furthermore, this connection between the risk factors of breast cancer and the use of Hormonal replacement therapy has been investigated for decades in many epidemiological studies. 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 breast cancer risk (Hamilton and Piccart, 2000). Smoking and alcohol consumption are also inferred to increase the risk of breast cancer (Gaudet *et al*., 2013).

1.5. Diagnostic tools for breast cancer

1.5.1. 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 ACS. However, there is no evidence on the effect of detection of breast cancer during breast self–examination, because it is highly inaccurate (Petro-Nustas *et al*., 2013).

1.5.2. Mammography

Mammography is a screening method for superficial diagnosis of breast cancer especially for women between the age of 39 and 69 years (Nalejska *et al*., 2014). In spite of the fact that mammography in combination with superior diagnostic methods has been reported to decrease mortality rates associated with breast cancer, it has a limited impact due to poor sensitivity of the technique to detect breast cancer (Toyoda and Ishikawa, 2011).

1.5.3. Ultrasound

Ultrasound is more sensitive than a mammogram especially in patients with dense breast tissue. In developed countries, like the United State, ultrasounds are becoming popular among lower level health centres due to its availability, inexpensive nature and non invasive characteristics (Leonard, 2003).

The limitations of ultrasound are that most of the cancers cannot be detected *via* an ultrasound because it only detects an abnormality, but it won't differentiate between cancers or benign conditions. Therefore additional procedures are required for differential diagnosis (Zhou, 2013).

1.5.4. Magnetic resonance imaging

Magnetic resonance imaging is commonly used to assess the presence of complex lesions in women with a high risk of developing breast cancer (Zhou, 2013). As a screening tool for breast cancer, it has shown a higher degree of sensitivity compared to mammography. However, the technology renders certain limitations like presence of false-negative and false-positive results, human error and / or patient's characteristics (França *et al*., 2017).

1.5.5. Molecular diagnostic methods

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 (Bhatt *et al*.,

2010). Molecular biomarkers serve as a reliable indicator of a biological state, behaviour and function of the cells (Yang *et al*., 2009). There are different molecular diagnostic methods which can be used to evaluate the presence of such biomarkers in a patient sample. These method includes techniques such as Quantitative Real Time Polymerase Chain Reaction (qRT – PCR), Enzyme–Linked Immunosorbent Assay, Fluorescent *in situ* Hybridization (FISH) and Immunohistochemistry (IHC) Assay (Nalejska *et al*., 2014).

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 (Amatori *et al*., 2017). 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 (Amatori *et al*., 2017).

FISH is a cytogenetic technique used to detect the physical location of a specific gene or nucleic acid sequences DNA or ribonucleic acid (RNA) in intact chromosomes and to assess if multiple copies of disease specific genes or nucleic acid sequences (e.g., HER-2 gene of breast cancer) are present. In a breast cancer patient, this technique can differentiate between malignant and pigment lesions (Press et al., 2016a).

IHC has paramount significance in the medical field, especially for pathological disease diagnosis. It is a standard technique used to detect the expression of disease specific proteins using an antibody. IHC is commonly used to evaluate HER-2 status in breast cancer patients (Zhao *et al*., 2014).

1.6. Breast cancer treatment

Breast cancer is a hormone-dependent tumor and estrogen is known to play a major role in the initiation and progression of the disease (Chang, 2012). Therefore, breast cancer therapy involves a multidisciplinary approach comprising surgery, radiotherapy, and neoadjuvant and adjuvant therapy. Effective therapy of breast cancer requires maximum therapeutic efficacy, with minimal undesirable effects to ensure a good quality of life for patients. There are two ways to treat breast cancer, depending on the type and stage of the cancer (Balaji *et al*., 2016a).

1.6.1. Local treatments

Treating the tumour without affecting the rest of the body. Types of local therapy used for breast cancer treatment includes: surgery and radiation therapy with these treatments being more likely to be useful for earlier stage cancers, although they might also be used in some other cases (Chang, 2012).

1.6.2. Systemic treatments

The second type of treatment is systemic therapies that can reach cancer cells anywhere in the body by using drugs which are administered orally or directly into the bloodstream. These include; chemotherapy, hormone therapy and targeted therapy (Leonard, 2003). Although many therapies have been designed against breast cancer, still the treatment of breast cancer continues to be a challenge due to an acquisition of drug resistance and relapse of the tumour (Chang, 2012). However, in this present study, the focus is on chemotherapy drugs specifically Tamoxifen treatment.

1.6.2.1. Chemotherapy treatments

The use of chemotherapy in cancer treatment started at the beginning of the 20th century with the aim of creating techniques to screen synthetic chemical compounds to be utilized on transplantable tumors in rodents after a surgical intervention and radiotherapy become common place in the field of cancer therapy and treatment (DeVita and Chu, 2008; Waks and Winer, 2019).

Chemotherapy is a remarkably different approach; it uses chemical agents such as anti-cancer or cytotoxic drugs to interact with cancer cells in order to completely destroy or control the growth of cancer cells. Instead of physically removing a tumor or a part of it when applying surgery or radiation therapy for treating cancer (El-Awady *et al*., 2017).

Chemotherapy is a systemic method of cancer treatment whereby the drugs are able to reach most parts of the body. Therefore, chemotherapy is likely to be recommended for cancer that has already spread to other areas of the body, for tumors that occur at more than one site, or for tumors that cannot be removed surgically. It is also used when a patient has recurrent disease after initial treatment with surgery or radiation therapy (Kolesnikova *et al*., 2019).

For some cancers, chemotherapy as primary treatment, can destroy all the cancer cells and cure the cancer. As a primary treatment, chemotherapy is used for some cancers such as Hodgkin's disease, leukemia, Burkitt's lymphoma, localized diffuse large cell lymphoma, Wilms' tumor, small cell lung cancer, and testicular cancer (Kolesnikova *et al*., 2019).

As an adjuvant treatment, chemotherapy is given prior to, or after other treatment methods, to increase the effectiveness of cancer treatment. Most often, adjuvant chemotherapy is given after other therapies have destroyed the clinically detectable cancer cells (Fraguas-Sánchez *et al*., 2019). The role of adjuvant chemotherapy is to reduce the risk of recurrence or to prolong survival. If total treatment is not possible, chemotherapy may be given to minimize the discomfort caused by cancer or slow the progression of the disease to prolong the patient's life (Hlaváč *et al*., 2013).

Chemotherapy may be given prior to surgical resection or radiation therapy to shrink the tumor and make it easier to resect. This type of chemotherapy is called neo-adjuvant, induction, or preoperative chemotherapy (Waks and Winer, 2019). As a palliative therapy, chemotherapy can be used in order to help make the cancer patient's life as comfortable as possible. In the case of Waldenstrom's macroglobulinemia, which is generally considered incurable, chemotherapy is administered to relieve symptoms and serious complications such as anaemia (Hlavata *et al*., 2012; Waks and Winer, 2019).

1.6.3. Tamoxifen: as a chemotherapy treatment

Tamoxifen has been used in the management of breast cancer for over 30 years. Since its introduction for the treatment of advanced breast cancer, its indications have increased to include the treatment of early breast cancer, DCIS, and more recently for breast cancer chemoprevention (Clemons *et al*., 2002).

Tamoxifen has a good tolerability profile and moreover, unlike many other endocrine therapies, it is efficacious in both pre- and postmenopausal women (Tabarestani *et al*., 2014). It is the combination of efficacy and tolerability that allows Tamoxifen to maintain its position as the hormonal treatment of choice for most patients with ER breast cancer (Chang, 2012).

The efficacy and safety of Tamoxifen monotherapy in advanced breast cancer has been compared with those of a number of other endocrine treatments, including androgens (fluoxymesterone),

progestins (medroxyprogesterone acetate and megestrol acetate), oestrogens (ethinyl oestradiol and diethylstilboestrol), other anti-oestrogens (toremifene), and various forms of ablative surgery (oophorectomy, adrenalectomy). While similar objective responses were achieved in both groups of patients, Tamoxifen was generally better tolerated (Criscitiello *et al*., 2011).

Tamoxifen has been widely used for the treatment and prevention of recurrence in a patient with estrogen receptor (ER) or progesterone receptor (PR) breast cancers (Callaghan and Higgins, 1995). Tamoxifen is the first line of treatment for breast cancer in South Africa because it reduces the rate of breast cancer resumption by approximately a half (Chang, 2012). Previous studies have shown that Tamoxifen is a substrate for ABC drug transporters (Eckford and Sharom, 2009). ABC drug transporters in breast cancer treatment appears to be associated with poor clinical outcomes in breast cancer patients (Mao and Unadkat, 2005).

1.6.3.1. Chemotherapy resistance

The management of cancer disease involves processes, which include surgery, radiotherapy and chemotherapy treatments. However, chemotherapy is usually an effective treatment for cancer patients with advanced and metastatic tumors or hematological malignancies (DeVita and Chu, 2008). Moreover, cancer cells often develop simultaneous resistance to many functionally and structurally unrelated anti-cancer drugs, due to a phenomenon known as multidrug resistance (MDR), which is a major problem in the treatment of cancer. Cancer cells with the MDR phenotype may have either inherent resistance to anti-cancer drugs or resistance acquired after cycles of chemotherapy (Waks and Winer, 2019).

The development of chemoresistance is a persistent and continuous problem during the treatment of local and disseminated disease. Cancers have the ability to develop resistance to traditional therapies, and increased prevalence of these drug resistant cancers necessitates further research and treatment development (Fraguas-Sánchez *et al*., 2019).

Drug resistance is a well-known phenomenon that results when diseases become tolerant to pharmaceutical treatments. This concept was first considered when bacteria became resistant to certain antibiotics, but since then similar mechanisms have been found to occur in other diseases, including cancer (Fletcher *et al*., 2016). Some methods of drug resistance are disease specific, while others, such as drug efflux, which is observed in microbes and human drug resistant cancers,

are evolutionarily conserved. Although many types of cancers are initially susceptible to chemotherapy, over time they can develop resistance through these and other mechanisms, such as DNA mutations and metabolic changes that promote drug inhibition and degradation (Karvar, 2014).

The main clinical problem in cancer treatment is the development of MDR, which is the determinant of chemotherapy failure. Therefore it's a critical issue in the management of breast cancer patients who use Tamoxifen as a chemotherapy (Fletcher *et al*., 2016; Hazlehurst and Hacker, 2009; Nakanishi, 2007; Y. Zhou *et al*., 2013).

Resistance constitutes a lack of response to drug-induced tumour growth inhibition; it may be inherent in a subpopulation of heterogeneous cancer cells or be acquired as a cellular response to drug exposure taking into consideration that the resistance varies from one patient to another. Tumors can be intrinsically drug resistant or develop resistance to chemotherapy (Karvar, 2014). MDR can develop by diverse mechanisms shown in Table 1.1.

Table 1.1: Multidrug resistance mechanisms.

Decreasing the rate of drug uptake.
Increasing the drug efflux (increase expression of ABC transporters members).
Alterations in the drug metabolism process.
Mutation of drug targets.
Activation of DNA repair mechanisms.
Evasion of apoptosis.

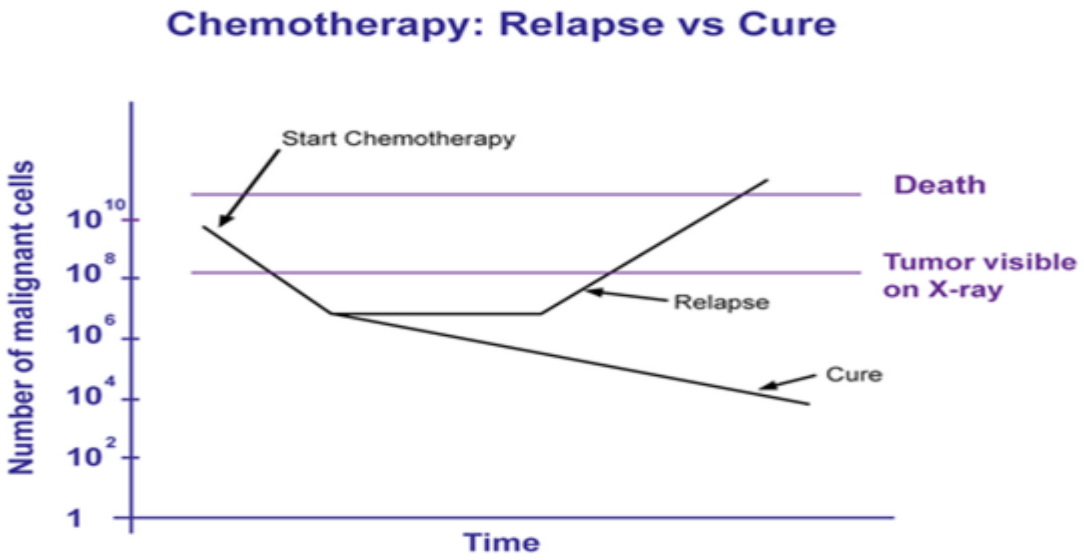

Figure 1.7: Represents the chemotherapy resistance diagram showing the time guideline in the chemotherapy course of treatment (https://chemoth.com/resistance).

The use of modern genomics, proteomics, bioinformatics, and systems biology approaches has resulted in a substantial increase in our ability to identify molecular mechanisms that are involved in MDR in cancer and to find drugs that may block or reverse the development of drug resistance (Karvar, 2014). Considerable attention has been paid to the role played by membrane transport proteins belongings to ABC transporters such as ABCB1 (Sharom, 2008) and ABCC1 (Sun *et al*., 2012).

1.7. Drug transporters

Membrane transporters proteins are important in order to preserve the organismal and cellular hemostasis and work on importing the essential nutrients for the cellular metabolism process and therefore exporting waste products and toxic compounds (Shu *et al*., 2003). Furthermore, transport proteins/drug transporters play a significant role in absorption, distribution, and excretion of many anti-cancer medications, as well as playing a crucial role in drug response serving as drug targets and determining the most effective dose of each drug (Leabman *et al*., 2003).

Based on function, membrane transporter proteins are divided into major super families which are the ABC and the solute carrier (SLC) transporters (Benjeddou, 2010). While the human ABC

transporter family comprises seven sub-families with 48 members, which include MDR1 protein, responsible for the pumping of xenobiotics from cells (Leabman *et al*., 2003), the SLC family has 47 subfamilies with more than 365 members that are responsible for the uptake and importing of neurotransmitters, nutrients, heavy metals and other substrates into cells (Benjeddou, 2010; Leabman *et al*., 2003).

1.7.1. SLC transporters

The SLC transporters are a large family of trans-membrane proteins comprised of 365 members and then subdivided into 47 families. These transporters can be found in most tissues, but primarily expressed in the liver, lungs, kidneys and intestines. Where they are either localized at the basolateral or apical plasma membrane of polarized cells as well as can be expressed in mitochondria and other organelles (Benjeddou, 2010; Wojtal *et al*., 2009).

Since they are membrane-associated transporters, they are crucial for facilitating the process of passing the solutes such as peptide, amino acids, bile acids, xenobiotic, ions, drugs and other biological compounds (Wojtal *et al*., 2009). Members of the SLC family act as influx transporters for nutrients and substances essential for the cell, thus they mediate main physiological functions and efflux of endogenous substrates including amino acids, lipids and bile acids, thyroid hormones and xenobiotics (Brockmöller and Tzvetkov, 2008).

Based on function and substrate specificity, SLC family divided into three subgroups: Organic Cation Transporters, the organic cation/zwitterion transporters, and the organic anion transports (Koepsell *et al*., 2007;Roth *et al*., 2012).

1.7.2. ABC drug transporters as a therapeutic tool

ABC drug transporters are a group of proteins that are located in the kidney, liver and intestinal cells which is considered useful therapeutically because of their native efflux activity (Sun *et al*., 2012). ABC drug transporters are members of a transport system superfamily that is one of the largest and is possibly one of the oldest families with representatives in all extant phyla from prokaryotes to humans, but recently has become prominent because of their essential target functioning to the chemotherapeutic response (Doyle and Ross, 2003).

The ABC transporter genes participate primarily in effluxing or pumping out endogenous and anti-cancer drugs from cells and any other metabolites (Shukla *et al*., 2011). ABC transporters are thus crucial in the management of chemotherapy resistance as effectors of treatment (Leonard, 2003). ABCB1 drug transport was the second ABC protein to be identified as a cause of MDR1. More recently, Breast Cancer Resistance Proteins (BCRP) (BCRP, MXR, ABCP, ABCG2) were identified in addition to other ABC proteins that have been initially identified for their physiological function in breast cancer (Nakanishi, 2007).

Increased expression of ABC family transporters are associated with chemotherapy failure (Nakanishi, 2007). Human BCRP has been shown to mediate drug resistance through energy-dependent efflux of drug substrates without the need for glutathione (Maliepaard *et al*., 2001). Thus, ABC drug transporters are one of the most valuable therapeutic biomarker targets in chemotherapy treatment. Breast cancer as disease represents one of the best models to study the functioning of these transporters in the effluxing of chemotherapy (Tamoxifen) leading to ineffective therapeutic intervention in breast cancer leading to poor prognosis and increased mortality rates (Karvar, 2014).

Table 1.2: ABC transporters family and their sub-families.

ABC SUBFAMILY ABCA	ABC SUBFAMILY ABCB	ABC SUBFAMILY ABCC	ABC SUBFAMILY ABCD	ABC SUBFAMILY ABCE	ABC SUBFAMILY ABCF	ABC SUBFAMILY ABCG
ABCA1	ABCB1	ABCC1	ABCD1	ABCE1	ABCF1	ABCG1
ABCA2	ABCB2	ABCC2	ABCD2		ABCF2	ABCG2
ABCA3	ABCB3	ABCC3	ABCD3		ABCF3	ABCG4
ABCA4	ABCB4	ABCC4	ABCD4			ACBG5
ABCA5	ABCB5	ABCC5				ABCG8
ABCA6	ABCB6	ABCC6				
ABCA7	ABCB7	ABCC7				
ABCA8	ABCB8	ABCC8				
ABCA9	ABCB9	ABCC9				
ABCA10	ABCB10	ABCC10				
ABCA11	ABCB11	ABCC11				
ABCA12		ABCC12				

1.7.2.1. ABC drug transporter structure

ABC proteins comprise one of the largest protein families, and members of these family are found in all living organisms from microbes to humans. The native spread and presence of these proteins with a relatively conserved structure and function put them forward for consideration to play fundamental roles (Jones and George, 2004). The basic structure that defines the members of these protein family are the combination of retained ATP-binding and transmembrane domains (TMDs). In mammals the functionally active ABC proteins consist of four characteristic domains, two TMDs, and two cytosolic ABC. Also commonly known as two nucleotide binding domains (NBDs) (Jones and George, 2004). The TMDs contain multiple hydrophobic segments, which span the membrane and form the transmembrane (TM) channel (Sharom, 2008).

The primary sequences of ABC transporter TMDs are significantly variable compared with those of the NBDs, which consists of the short motifs involving ATP-binding the Walker A Motif, also known as the P-loop. Structurally, the Walker A Motif consists of an α-helix and is always followed by a glycine-rich loop and also, in addition the site has a primary amino acid sequence of GXXGXFKS(or T) (Jones and George, 2004; Sharom, 2008). The letter X can represent any amino acid. Walker B Motif, is a β-strand. The Walker Motifs are connected to each other by a peptide sequence of about 100 residues. Structurally, these connecting residues fold into an α-helical domain, with the primary amino acid sequence of the site is being hhhhD, in which h represents any hydrophobic amino acid. The C Motif, also known as the Signature motif, LSGGQ motif, or the linker peptide, or known as the diagnostic signature sequence of ABC proteins, directly following the Walker B Motif, and has a primary amino acid sequence of LSGGQQ/R/KQR. The switch motif has been found to be located at the end of the β4-strand in ATP binding proteins (Jones and George, 2004; Sharom, 2008).

Due to the variety of different amino acids that can be used in the primary sequence, of both the Walker Site A and B, the non-variant amino acids within the sequence are highly conserved. A mutation of any of these amino acids will affect the binding of ATP or interfere with the catalytic activity of the enzyme (Domenichini *et al*., 2019). The primary amino acid sequence determines the three dimensional structure of each motif. While the TMDs form the TM channel and are thought to contain the substrate binding sites, the NBDs are molecular motors that transform the chemical potential energy of ATP into protein conformational changes. These four domains could

be present within one polypeptide chain, which are called full transporters, or within two separate proteins, which are called half transporters. All of the ATP binding domains are made up of an estimated 250 residues and two subunits, creating a dimer. These residues are folded into six α-helices and five β-strands (Domenichini *et al*., 2019).

Eukaryotes only have export ABC transporters, which commonly have a single polypeptide for the core structure with each NBD being C-terminal to each TMD. There are exceptions to this scheme, with the most usual of these being half-transporters found in both prokaryotes and eukaryotes, in which each TMD is fused covalently to a C-terminal NBD, or with an N-terminal NBD followed by a TMD (Robey *et al*., 2018).

The complete core structures for half-transporters can be homo or heterodimers. Therefore, the ABC transporters become functionally competent after specific dimerization. The membrane-spanning domains form the key structural background of the function of ABC transporters. The two TMDs contain polypeptide chains that span the membrane multiple times, typically forming six TM a- helices per domain, a total of twelve helices in a full transporter (Domenichini *et al*., 2019).

The 12 TM a-helices probably form a pore-like structure across the membrane, and through this path a range of different substrates can be transported by these proteins. The conformational changes within the TMD domains are believed to be responsible for the opened or closed states of these TM structures. TMD domains are believed to be responsible for the opened or closed states of these TM structures (Domenichini *et al*., 2019; Robey *et al*., 2018).

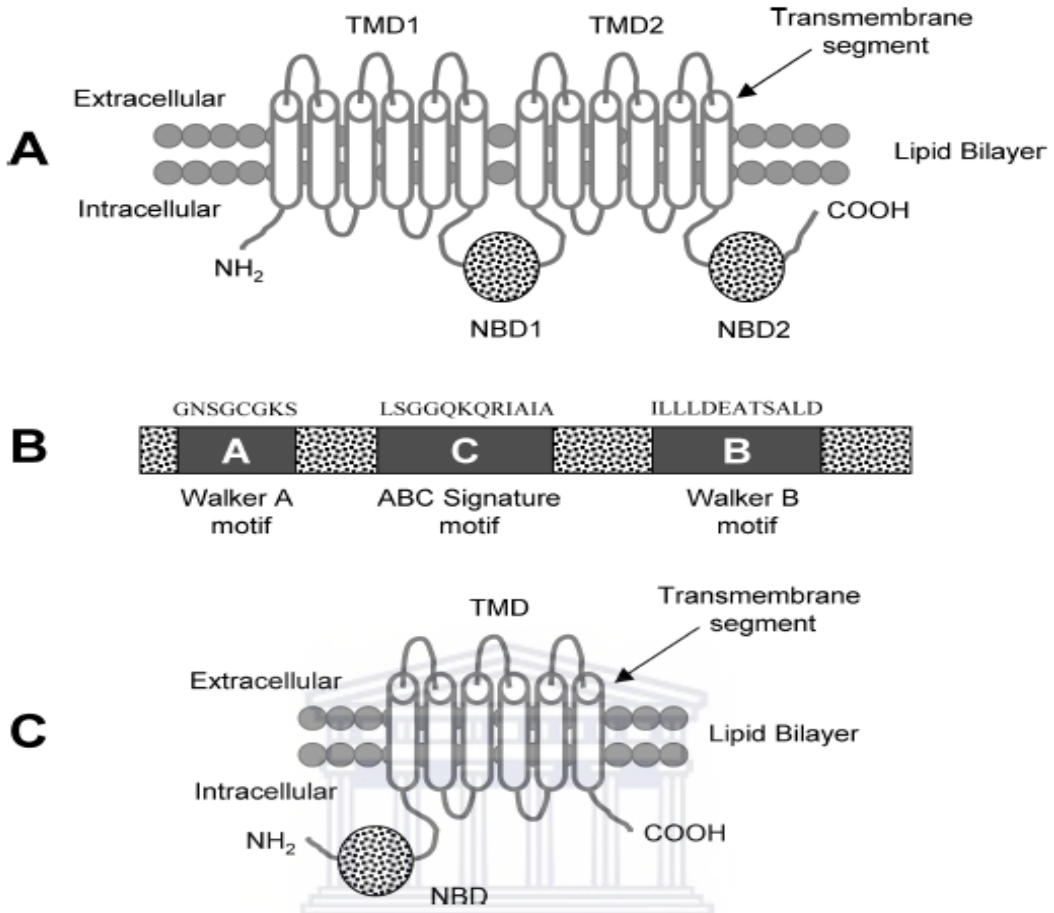

Figure 1.8: Represents the typical ABC transporter structure. A, B, and C represents the Walker A Motif, Walker B Motif, and the C Motif respectively.

1.7.2.2. ABC drug transporter mechanism of action

The transport of organic and inorganic molecules across cellular membranes is vital to all forms of life, as it allows cells to maintain an off equilibrium condition. ABC proteins can utilize the energy derived from ATP hydrolysis to perform a directed TM movement of their substrates (primary active transporters), open or close a specific membrane channel (e.g. ion-channels), or regulate the permeability of multi-protein channel complexes (receptors). In the ABC proteins, which act as primary active transporters, the transport function depends on the hydrolysis of ATP within the NBDs (Domenichini *et al*., 2019).

These cytoplasmic domains are attached to the intracellular regions of the TMDs, and a close interaction provides the functional connection between these two different domains. The nucleotide-binding domains bind cytoplasmic ATP and, in the active transporters, ATP hydrolysis ensures the energy for the uphill transport of a substrate (Domenichini *et al*., 2019; Robey *et al*., 2018). The specific close interaction of NBDs with the TMDs provide the transmission gear of the conformational changes caused by substrate binding and the hydrolysis of ATP. The NBDs contribute to the transmission interface primarily through amino acid side chains around their Q-loops. Extraordinarily, the TMDs of these transporters, despite their vastly different topologies and architectures, all feature similar alpha helices that provide the bulk of the contacts to the NBDs (Domenichini *et al*., 2019; Robey *et al*., 2018).

Each ATP binding motif has a different role to play whether it is directly involved with the binding of ATP or helping with the construction of the ABC transporter (El-Awady *et al*., 2017). The ATP molecule binds to the connecting point of each subunit of the dimer, indicating that ATP is in close proximity to both subunits during catalysis. The two binding motifs that ATP directly interacts with is the residues from the Walker A Motif, located on one of the subunits, and the residues from the C binding motif, located on the other subunit. The Walker A binding motif has a lysine side chain, which is essential for the binding of ATP (Chen *et al*., 2016). The lysine residue forms hydrogen bonds with the oxygen atoms of two phosphate groups within ATP, therefore creating proximity and orientation of ATP in the binding site (Kathawala *et al*., 2015).

In order for the Walker A Motif to bind to ATP, the ATP molecule must be in the binding site. The signature motif acts as a signal to the Walker A Motif, letting the Walker A know when the ATP molecule has bound to the binding site. The Signature motif does this by allowing its residues to extend from the subunit they are located on into the other subunit, where the Walker A Motif is (Chen *et al*., 2016). It is necessary that ATP binds to both NBDs in order to complete the catalytically active structure. The Walker B Motif contains the amino acid glutamate within the short sequence. Glutamate can be used to perform a nucleophilic attack on the ATP molecule (Fletcher *et al*., 2016).

Found in the switch binding motif is a histidine residue. The function of the histidine is to influence the reaction catalytically by contacting the residues across the dimer interface, including the

Walker A Motif and the Walker B Motif. It is the histidine residue that forms the tight coupling between the binding of the ATP molecule and the dimer (Fletcher *et al*., 2016).

Following the hydrolysis of adenosine diphosphate, a conformational change must occur to separate the ATP-binding cassette. This separation is driven by an electrostatic repulsion by the adenosine diphosphate product that is bound to the Walker A Motif with the inorganic phosphate product bound to The C Motif (Li *et al*., 2016).

1.8. Biomarkers

In the recent years, knowledge about cancer biomarkers has enormously increased providing great opportunities for improving the management of cancer patients by enhancing the detection efficiency and treatment efficacy (Nalejska *et al*., 2014). Recent technological advancements have enabled the examination of many potential biomarkers and renewed interest in developing new biomarkers.

Biomarkers of cancer could include a broad range of biochemical entities, such as nucleic acids, proteins, sugars, lipids, and small metabolites, cytogenetic and cytokinetic parameters as well as whole tumour cells found in the body fluid (Chatterjee and Zetter, 2005; Nalejska *et al*., 2014).

A comprehensive understanding of the relevance of each biomarker will be very important not only for diagnosing the disease reliably, but also aid in the choice of multiple therapeutic alternatives currently available that is likely to benefit the patient. Various biomarkers for diagnosis, prognosis and therapeutic purposes, which include markers already in clinical practice as well as various upcoming biomarkers are being studied (Bhatt *et al*., 2010; Colomer *et al*., 2018).

1.8.1. Cancer biomarkers

Biomarkers have a very important role in disease diagnosis, predictive and therapeutic treatment outcomes. While biomarkers help in differentiating physiological and pathological mechanisms, they are equally important in assessing disease response to the medications therapeutic outcome, disease progression and to explore disease mechanisms (Strimbu *et al*., 2010).

Moreover, 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.

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 determine the patients who are at a high risk of disease development at a later stage (Bhatt *et al*., 2010).

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. In breast cancer, BRCA1, HER-2/neu, ER and PR are a few of the important prognostic biomarkers (Bhatt *et al*., 2010; Duffy *et al*., 2017; Weigel and Dowsett, 2010).

Predictive biomarkers are used to determine relapse or the recurrence of disease after a patient had undergone treatment such as the surgical removal of tumours. ER, PR, and HER2 could serve as important predictive biomarkers to envisage sensitivity to endocrine therapy and Herceptin treatment in breast cancer respectively (Bhatt *et al*., 2010).

1.8.1.1. Breast cancer biomarkers

Breast cancer treatment has been going through several changes in the past decades as a consequence of the technological advancements in oncology clinical practices and the innovative discovery of specific prognostic and predictive biomarkers that enable the application of more individualized therapies to different breast cancer molecular subgroups (Liu *et al*., 2015). Hence, these subgroups show specific differences regarding biological clinicopathological behavior and features (Weigel and Dowsett, 2010). In addition to the classical clinical prognostic and predictive factors of breast cancer, biomarkers have played an essential and significant role in the management of patients with invasive breast cancer (Nalejska *et al*., 2014). Biomarkers aid in selection of patients who likely would respond to systemic therapy, such as both the molecular biomarkers; ER and PR that both have played important roles in patient selection for systemic treatment response and should be measured on all newly diagnosed invasive breast cancers (Duffy *et al*., 2017).

Biomarker investigation, examination, and analysis in all cancers, not only breast cancer, provides additional information about classical clinical factors, but also enables patients, with a more favorable benefit or/and risk balance, to receive certain treatments with the best possible beneficial clinical outcome (Colomer *et al*., 2018). In breast cancer, biomarker analysis in the oncology clinical practices is a regular procedure that originally began with testing for hormone receptor expression to guide Tamoxifen therapy (Colomer *et al*., 2018; Tabarestani *et al*., 2014). Although several breast cancer biomarkers have been developed, their predictive value in diagnosis and detection of recurrence of the disease is a controversial matter (Weigel and Dowsett, 2010).

1.8.1.1.1. Genetic markers

Breast cancer type 2 susceptibility protein (BRCA2) locus is the most studied genetic biomarker in breast cancer. BRCA2 is a tumour suppressor gene and is involved in DNA repair pathways such as homologous recombination (Sharan *et al*., 1995; Stefansson *et al*., 2009). The expression of BRCA2 is observed in different cancer tissues such as prostate, cervix and ovarian cancers. 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 Recombinase and Partner and Localizer of BRCA2 Failure to repair DNA damage can lead to replication errors resulting in DNA mutations and cancer (Petrucelli *et al*., 2019; Stefansson *et al*., 2009). Lowry *et al*. (2012) have reported that amongst the hundreds of mutations in the BRCA2 gene, several of its mutations have the potential to initiate breast cancer.

1.8.1.1.2. 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 biomarker for breast cancer using the FISH technique (Press *et al*., 2016). 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 (Press *et al*., 2016; Sjöblom *et al*., 2006). It is reported to be overexpressed in approximately 25-30 % of breast cancer cases and its expression is particularly high in invasive breast cancer specimens. Furthermore, overexpression of HER-2 has been associated with the negative expression of ER. HER-2 is an important biomarker target for the treatment of the disease (Mirtavoos-Mahyari *et al*., 2014).

1.8.1.1.3. Serum markers

Cancer antigen (CA 15-3) is one of the most extensively studied circulating prognostic factors for breast cancer. Preoperative concentrations in combination with existing prognostic factors are useful in predicting disease outcomes in patients with newly diagnosed breast cancer (Chatterjee and Zetter, 2005). Cancer antigen (CA 27- 29), however, is a less widely used serum marker in breast cancer (Duffy *et al*., 2017). Proteins in the MUC-1 family which includes CA 15-3, 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 cancer-associated antigen MUC1 is not likely to be advantageous (Donepudi *et al*., 2014).

1.8.2. ABC drug transporter as a therapeutic target

MDR is a key determinant of cancer chemotherapy failure. One of the major causes of MDR is the enhanced efflux of drugs by membrane ABC drug transporters. Targeting ABC transporters is a promising approach for eliminating or suppressing drug resistance in cancer treatment (Choudhuri and Klaassen, 2006; Kolesnikova *et al*., 2019; Moriya *et al*., 2002).

In cancer cells, the leading cause of MDR is the increased efflux of anti-cancer drugs by increasing the expression of membrane embedded drug transporters. The most well-studied drug transporters in humans is (P-glycoprotein/MDR1/ABCB1), which belongs to the ABC transporters B subfamily (El-Awady *et al*., 2017; Suresh *et al*., 2015). Other well studied ABC drug transporters include MRP1/ABCC1, breast cancer resistant protein (BCRP/MXR/ABCP/ABCG2).

Clinical evidence reveals that >90% of metastatic cancer patients face treatment failure due to drug resistance phenomena (Estudante *et al*., 2016). Therefore, MDR creates a major challenge to effective and successful chemotherapy in cancer treatment. As significant contributors that render cancer cell drug resistance, ABC proteins belong to one of the largest protein superfamily that give researchers hope to unravel the molecular mechanism of drug resistance and potentially a therapeutic biomarker target for successful chemotherapy treatment of breast cancer (Domenichini *et al*., 2019).

1.9. Methods for biomarkers discovery

1.9.1. 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. breast cancer). In breast cancer, useful biomarkers for breast cancer diagnosis have been identified during such studies. Examples identified by this technique are Serum biomarkers to detect breast cancer (BC1, BC2, BC3, and CA 15-3) (Baak *et al*., 2005; Khalilpour *et al*., 2017).

1.9.2. Transcriptomics

Genetic profiling has been used in medical research for the identification of genetic markers required for disease screening including cancer (Ginsburg and Willard, 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 messenger ribonucleic acid (mRNA) expression levels of thousands of genes in an organism (Karakach *et al*., 2010; Colomer *et al*., 2018; Ginsburg and Willard, 2009).

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 complementary deoxyribonucleic acid (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 (Karakach *et al*., 2010).

DNA microarray analysis involves 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 (breast cancer cells or tissue) samples, (c) transcription of extracted mRNA into cDNA molecules that are differently labelled (the controlled sample is labelled with CS) 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).

1.9.3. 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, 2012). It is concerned with storage, extracting, organizing, interpreting and utilizing information from biological sequences and molecules (Baxevanis and Francis Ouellette, 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 (Choudhuri, 2014; Khalid Raza, 2012; Kunin *et al*., 2008). 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 (Choudhuri, 2014).

With advancements in computational technology, bioinformatics continues 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, 2012). Furthermore, bioinformatics has presented ways in which data mining approaches can be used to filter valuable targets such as microRNAs (miRNA), genes, or proteins for the discovery of possible novel biomarkers for diseases (Liang, 2003).

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 signaling, proliferation, communication and specificity of disease metabolism is through cancer bioinformatics (Choudhuri, 2014; Liang, 2003). 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. 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 (Raza, 2012).

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 (Yang *et al*., 2009).

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 (Liang, 2003; Wang *et al*., 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.

1.10. Aim of the study

The aim of this research was to study the prominent ABC transporters involved with Tamoxifen MDR in breast cancer treatment and management. Using several *in silico* methodologies.

Objectives

1. Categorize and prioritize ABC transporter genes *via* data and text mining.
2. To perform an *in silico* tissue specific expression analysis of the shortlisted ABC targeted genes using TiGER and HPA Databases.
3. To perform functional network interactions and expansion of the prioritized ABC transporter genes *via* STRING.
4. Intersectional venn diagram to identify toxicity relation between the shortlisted genes *via* CTD.

5. Survival and Predictive analysis of ABC prioritized genes for breast cancer using PROGGENE databases.
6. Prognostic/ predictive validation of ABC prioritized genes using datasets from SurvExpress and PROGGENE databases.

CHAPTER TWO: Categorization, prioritization through gene expression and functional network interaction analysis of the ABC transporter genes implicated in Tamoxifen resistance

2.1. Introduction

Breast cancer is a global health problem that continues to affect women and one of the leading causes of death in most cases. In South Africa, an estimated age standardized incidence rate of 128 per 100000 women in 2017, and age standardized mortality rate of 113 per 100000 women were reported (NCR, 2017).

Breast cancer is a hormone-dependent tumour, and estrogen is known to play a major role in the initiation and progression of the disease (Chang, 2012). Increasing interest in personalized medicine has led to in depth research into genetic pathways of drug metabolism and the role of biomarkers to optimize therapeutic decisions for each individual patient. Biomarkers has received considerable attention from the research community (Massarweh *et al.*, 2008; Bhatt *et al.*, 2010; Nalejska *et al.*, 2014). Therefore, researchers from various fields, especially those involved in biomedical research aim to find genetic biomarkers indicative of breast cancer. Novel biomarkers can be clarified from the existing literature. However, the huge amount of scientific publications on breast cancer makes this a challenging task (Bhatt *et al*., 2010).

Tamoxifen is the first line of treatment of breast cancer in South Africa but has been widely used for the treatment and prevention of recurrence in patients (Callaghan and Higgins, 1995; Chang, 2012). Tamoxifen reduces the rate of breast cancer resumption by approximately half. Previous studies have shown that Tamoxifen is a substrate for ABC drug transporters (Tabarestani *et al*., 2014a). ABC drug transport in breast cancer treatment appears to be associated with poor clinical outcomes in breast cancer patients (H. Zhou *et al*., 2013).

Resistance to chemotherapy is a critical issue in the management of breast cancer patients who are treated with Tamoxifen (Nakanishi, 2007). Considerable attention has been paid to the role played by membrane transport proteins belonging to the ABC transporter family (Sharom, 2008). ABC drug transporters are a group of proteins that are located on the kidney, liver and intestinal cells considered useful therapeutically because of their native efflux activity.

They are members of a transport system super family that is one of the largest and is possibly one of the oldest families with representatives in all extant phyla from prokaryotes to humans, but recently has become prominent because of their essential target functioning to the chemotherapeutic response (Doyle and Ross, 2003).

They participate primarily in effluxing or pumping out endogenous and anti-cancer drugs from cells together with any other metabolites. ABC transporters are thus crucial in the management of chemotherapy resistance as effectors of treatment (Szakács *et al*., 2008). Breast cancer as disease represents one of the best models to study the functioning of these transporters in the effluxing of Tamoxifen leading to ineffective therapeutic intervention in breast cancer leading to poor prognosis and increased mortality rates (Cocca *et al*., 2016).

Many methods have been developed to identify new ways including biomarkers for diagnosis and therapeutic purposes for breast cancer (Frances, 2013). However, the research spans a wide range of techniques from wet-lab testing by biologists to computational methods by computer scientists. The latter research is promising because it helps reduce the number of molecules considered as potential biomarkers (Nalejska *et al*., 2014).

2.2. Data mining

The large available datasets raise its own problems like the presence of high level of redundancy found in these sequences. Therefore, data mining also known as knowledge discovery in databases, which involves mining of specific data of interest based on selective criteria from a vast amount of data plays a crucial role in data prioritization (Jurca *et al*., 2016). It could also be categorized as the extraction of information from different sources of published literatures (Yang *et al*., 2009).

In bioinformatics, data mining includes the optimization of disease treatment, the function of proteins as well as the prognostics, diagnostics, and therapeutics of diseases. Data mining has also been used to identify disease-associated entities and to understand their roles in disease onset as well as progression. Because of the unprecedented growth of genomic data, data mining has become a field of interest as almost any new discovery or development about a gene its pathways or protein it codes for, is recorded in literature and consequently updated in

databases. Thus the goal of data mining is to filter that knowledge and to present the resulting information to a user in a concise, understandable format (Jurca *et al*., 2016).

2.3. Text mining and useful databases

Text mining of available biomedical text, data, and information has greatly boosted target discovery in the 'omics' era. Target discovery is the key step in the biomarker and drug discovery pipeline to diagnose and fight human diseases (Stenson *et al*., 2012). Text mining can be defined as the computational discovery of new, previously unknown information, by automatically extracting information from different written resources (Debouck, 2009).

The usage of computational tools in information mining to unravel organic information has brought about new methods for biomarker and infection revelation and in addition, disease pathway explanation (Jurca *et al*., 2016). An *in silico* approach to target revelation for early conclusion, comprehension of diseases and improvement of therapeutics can spare impressive wet-lab time. This approach can uncover essential confirmation for transformative and practical connections amongst genes and proteins (Daugelaite *et al*., 2013).

2.3.1. Google Scholar

Google Scholar is a freely accessible website available at https://scholar.google.co.za/ that allows the search of a wide assortment of academic abstracts and literature. It draws on information from university repositories, journal publishers and other websites that has been identified as scholarly by Google Scholar (Baxevanis and Francis Ouellette, 2004).

In November 2013, a trademark 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 (Baxevanis and Francis Ouellette, 2004).

2.3.2. The National Center for Biotechnologies Information (NCBI)

The NCBI is part of the United States National Library of Medicine, a branch of the National Institutes of Health. The NCBI is a freely accessible website available at https://www.ncbi.nlm.nih.gov/. It houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatic tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for

biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through the Entrez search engine (Whirl-Carrillo *et al*., 2012).

2.3.2.1. PubMed

PubMed accessible at https://www.ncbi.nlm.nih.gov/pubmed is a free search engine, which accesses the MEDLINE database of references. PubMed, the free, home and office-based MEDLINE was first released in January 1996. In June 1997, the PubMed system was offered free to the public by the United states of America's Vice President, Al Gore when MEDLINE searches *via* the Web were established during a ceremony (Baxevanis and Francis Ouellette, 2004).

2.4. Databases for data mining

2.4.1. TiGER database

TiGER is a public database available freely at http://bioinfo.wilmer.jhu.edu/tiger/. The database was developed by the Bioinformatics Lab at the Wilmer Eye Institute at the Johns Hopkins University (Liu *et al*., 2008).

Understanding how genes are expressed and regulated in different tissues is a fundamental and challenging question. However, most of the current available biological databases do not focus on tissue-specific gene regulation (Liu *et al*., 2008). TiGER contains three types of data including tissue-specific gene expression profiles, combinatorial gene regulations, and cis-regulatory module (CRM) detections (Liu *et al*., 2008).

2.4.2. HPA atabased

HPA is a public database that provides protein expression profiles for a huge amount of normal and cancer tissue proteins from humans and are presented as immunohistochemistry images (Crowgey *et al*., 2018). In 2005, the first version of the HPA 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 (Prassas *et al*., 2012). It is a freely available interactive resource included as part of the HPA 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 (Baxevanis and Francis Ouellette, 2004).

The HPA contains numerous data on gene expression and alterations in the expression of these genes and can aid in the discovery of biomarkers for disease using an *in silico* approach (Crowgey *et al*., 2018; Prassas *et al*., 2012).

2.4.3. STRING Database

The STRING version 10 (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 (Yu *et al*., 2006). 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 (Prassas *et al*., 2012). The basis of these associations are through data mining of various databases and literature mining throughput, *in vitro* data as well as genomic context analysis prediction (Sun *et al*., 2005a).

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 (Sun *et al*., 2005a).

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 five million proteins, more than 200 million interactions and more than 1000 organisms (Daugelaite *et al*., 2013).

2.4.4. CTD Database

The CTD (http://ctdbase.org/) provides insights by curating and integrating data describing relationships between chemicals, genes/proteins, molecular pathways, gene ontology and phenotypes, as well as human diseases (Mattingly *et al*., 2003).

The etiology of many chronic diseases such as breast cancer involves interactions between environmental factors and genes that modulate physiological processes such as ABC drug transporter genes. Understanding interactions between environmental chemicals and genes/proteins may provide insights into the mechanism of a chemical's actions, disease susceptibility, toxicity, and therapeutic drug interactions (Davis *et al*., 2019).

The CTD provides curated and integrated toxicological data to facilitate hypotheses about chemical-gene interaction networks and the impact of chemical exposures on human diseases. Knowledge derived from CTD has the potential for predicting toxicity, identifying biomarkers of exposure, and unveiling putative therapeutic targets for diseases (Davis *et al*., 2019).

2.5. Aim and objectives:

This chapter aims to categorize and prioritize ABC drug transporter genes associated with breast cancer, chemotherapy resistance as well as explore predicted interactions and pathways in which the identified ABC transporter genes could be involved in multi drug resistance as well as their variability to chemotherapy treatment response to Tamoxifen.

The work outlined in this chapter was undertaken using several *in silico* methods.

The specific study objectives were to:

1) Identify all ABC drug transport genes using text mining.
2) Categorize and prioritize the list of ABC drug transporter genes identified through text and data mining.
3) Perform an *in silico* tissue specific expression analysis of the prioritized list of ABC transporter genes using the TiGER and HPA databases.
4) Perform functional network interaction of the prioritized list of ABC transporter genes *via* STRING.
5) Draw intersectional venn diagrams to identify toxicity relation between the prioritized genes *via* the CTD.

2.6. Methods and materials

2.6.1. Text mining and data extraction

The following platforms (Google Scholar, NCBI and PubMed) were launched using the following URLs (https://scholar.google.com/, https://www.ncbi.nlm.nih.gov/, and https://www.ncbi.nlm.nih.gov/pubmed/) respectively, to search for all ABC drug transporter family members in abstracts and journal articles implicated in breast cancer and their relation to chemotherapy resistance. Subsequent to these mining approaches, a final list of all ABC drug transporter genes were compiled.

2.6.2. Databases for data mining and extraction

2.6.2.1. Tissue-Specific Gene Expression Profiles *via* TiGER

The TiGER database was accessed at (http://bioinfo.wilmer.jhu.edu/tiger). On the home page, there are different experiment tool options which allow the user to view the expression of the genes of interest using Gene view, Transcription Factor View, and Tissue view. The ABC prioritized list gene ID's were used as input in the "search box" option in the Gene View tool to search for their expression profiles in breast tissue.

TIGER produce an outcome in three subsets of data based on three principles, (i) Expressed Sequence Tags (ESTs); (ii) (CRMs); and (iii) Combinatorial Gene Regulation. The EST module was selected for this study. Results were generated and displayed in EST plot form.

2.6.2.2. Protein expression profiling *via* HPA database

The HPA database was accessed at (https://www.proteinatlas.org). On the home page, several gene expression experiment tools are found including Tissue atlas, Cell atlas, Pathology atlas, Brain atlas, Blood atlas, Metabolic atlas as well as the general "search box" that provides an overall summary of all the above mentioned experiments. However, the general "search box" was used with the ABC gene IDs as input. Following this selection, the option "Tissue atlas" was followed by the selection of the "atlas tool" tab. This selection allowed for the focusing on a particular aspect of a genome-wide analysis of the human proteins i.e. protein-tissue expression profiling generated as immunohistochemistry-based maps in breast cancer tissue.

2.6.2.3. Gene/Protein Interaction Networks *via* STRING

STRING version 10 database was launched at (https://string-db.org) and Genes IDs for the ABC transporter genes identified in the Section 2.7.2, which are up-regulated in post treatment breast cancer were used as input for the generation of a gene-gene functional interaction network (Meiring, 2003; Franceschini *et al*., 2013). The ABC transporter genes were used as driver genes to produce expression networks. For the production of the functional and expression network, parameters were chosen as a follows: (i) a confidence level of 0.7, (ii) a network depth of the shortlisted genes and (iii) restricting to show only the interactions between the prioritized ABC transporter genes identified below in Section 2.7.2.

2.6.2.4. The CTD database

The CTD was launched at (http://ctdbase.org), and on the home page the "analyze" tab was used and "MyGeneVenn" option was selected to create a Venn diagram. The Venn diagram tool allows the user to create a comparison and identify the toxicity relationship between the chemicals and genes/proteins of interest. However, the NCBI symbols of the ABC prioritized genes of list 2 were used as input into the "search box" option in the "MyGeneVenn" tool. After which, the option" chemicals" were selected for the identification of interactions with the ABC prioritized genes of list 2 with "Tamoxifen" selected as the chemical of interest and the second box option left blank, as it is optional. The Venn diagram analysis was submitted and result outputs presented as below (Figure 2.5).

2.7. Results and discussion

2.7.1. Analysis of text mining and extraction results

In this section, several text-mining tools such as Google Scholar, NCBI, and PubMed were used. This was done in a bid to create a list of the ABC drug transporter genes involved with Tamoxifen resistance in the treatment of breast cancer.

The analysis revealed that 48 genes (List 1) belonging to this family of drug transporters are related to breast cancer. According to the literature and text mining of ABC drug transporters, based on their genetic structures, they are classified into seven families denoted from A to G (ABC**A**/ ABC**B**/ ABC**C**/ ABC**D**/ ABC**E**/ ABC**F**/ ABC**G**). The ABC**A** subfamily contains 12 members (ABCA1-ABCA12), ABC**B** subfamily contains 11 members (ABCB1-ABCB11),

ABC**C** subfamily contains 12 members (ABCC1-ABCC12), ABC**D** subfamily contains four members (ABCD1-ABCD4), ABC**E** subfamily only has only one member (ABCE1), ABC**F** subfamily contains three members (ABCF1- ABCF3), and the ABC**G** subfamily contains five members (ABCG1, ABCG2, ABCG4, ABCG5, and ABCG8).

The literature mining implicated 48 ABC drug transporter genes that are related to breast cancer. The 48 genes (List 1) were further prioritized based on the inclusive criteria Tamoxifen that all categorized genes are related to breast cancer, all categorized genes are related to Tamoxifen resistance, and all categorized genes are related to Tamoxifen post-treatment.

Prioritization of the ABC drug transporter genes according to the study demonstrated that some ABC transporter genes were significantly up regulated in breast cancer post treatment using Tamoxifen chemotherapy. The up-regulation of ABC transporter genes post treatment are of significance since the aim of the study was to investigate associations between the expression of ABC transporters and prognosis of breast carcinoma patient's treatment.

2.7.2. Prioritization of ABC shortlisted genes *via* inclusive selection criteria

List 1 of all 48 ABC transporter genes was further examined using text and data mining databases such as TiGER and HPA. The ABC transporter genes were used individually and the gene IDs used as input in the gene search box in both databases. Prioritization for the shortlisted genes were based on specific criteria which were: expression in breast tissue; expression in breast cancer tissue; role of gene; its function; involvement in drug transportation; and involvement in MDR to Tamoxifen. 10 of the 48 genes (from list 1) were able to satisfy the abovementioned criteria and were saved as List 2. The 10 genes (10 genes, ABCB2, ABCB9, ABCB10, ABCC1, ABCC4, ABCC5, ABCC10, ABCC11, ABCC12, ABCD1) from list 2 are discussed in detail in Table 2.1.

Table 2. 1: List of up-regulated genes in the chemotherapy post-treatment for breast cancer.

Gene Symbol	Full name of the gene	Expression in breast tissue	Involve in drug transportation	Involve in MDR	Role of the gene and its function
ABCB2	Transporter 2, ATP Binding Cassette Subfamily B Member (MDR)	YES	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance (Litviakov et al., 2016).
ABCB9	ATP Binding Cassette Subfamily B Member 9	YES	YES	YES	Involved in MDR as well as antigen presentation, translocation of peptides from the cytosol into the lysosomal lumen. Alternative splicing of this gene results in distinct isoforms which are likely to have different substrate specificities and transport various molecules across extra- and intra-cellular membranes (Ohashi-Kobayashi et al., 2006).
ABCB10	ATP Binding Cassette Subfamily B Member 10	YES	YES	YES	transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance (Sjöblom et al., 2006a).
ABCC1	ATP Binding Cassette Subfamily C Member 1	YES	YES	YES	Involved in multi-drug resistance, transports glucuronides and sulfate conjugates of steroid hormones and bile salts. Alternatively, spliced variants and transport various molecules across extra-and intra-cellular membranes. This protein functions as a multispecific organic anion

					transporter, with oxidized glutathione, cysteinyl leukotrienes, and activated aflatoxin B1 as substrates. This protein also transports glucuronides and sulfate conjugates of steroid hormones and bile salts (Sjöblom et al., 2006b).
ABCC4	ATP Binding Cassette Subfamily C Member 4	YES	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance. This family member plays a role in cellular detoxification as a pump for its substrate, organic anions. It may also function in prostaglandin-mediated cAMP signaling in ciliogenesis (Sjöblom et al., 2006b).
ABCC5	ATP Binding Cassette Subfamily C Member 5	YES	YES	YES	This protein functions in the cellular export of its substrate, cyclic nucleotides. Studies show that this protein provides resistance to anticancer drugs and acts as a multispecific organic anion pump which can transport nucleotide analogs. This protein functions in the cellular export of its substrate, cyclic nucleotides. This export contributes to the degradation of phosphodiesterases and possibly an elimination pathway for cyclic nucleotides. Studies show that this protein provides resistance to thiopurine anticancer drugs, 6-mercatopurine and thioguanine, and the anti-HIV drug 9-(2-phosphonylmethoxyethyl) adenine. This protein may be involved in resistance to thiopurines in

					acute lymphoblastic leukemia and antiretroviral nucleoside analogs in HIV-infected patients (Wijnholds et al., 2000).
ABCC10	ATP Binding Cassette Subfamily C Member 10	YES	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance (Hopper-Borge et al., 2004).
ABCC11	ATP Binding Cassette Subfamily C Member 11	YES highly expressed	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance. The product of this gene participates in physiological processes involving bile acids, conjugated steroids, and cyclic nucleotides. In addition, a single nucleotide polymorphism in this gene is responsible for determination of human earwax type. This gene and family member ABCC12 are determined to be derived by duplication and are both localized to chromosome 16q12.1 (Bera et al., 2001).
ABCC12	ATP Binding Cassette Subfamily C Member 12	YES	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance and Increased expression of this gene is associated with breast cancer. This gene is a member of the MRP subfamily which is involved in multi-drug resistance. This gene and another subfamily member are arranged head-to-tail on chromosome

					16q12.1. Increased expression of this gene is associated with breast cancer (Bera et al., 2002).
ABCD1	ATP Binding Cassette Subfamily D Member 1	YES	YES	YES	Transport various molecules across extra- and intra-cellular membranes and involved in MDR which is chemotherapy resistance. This protein is a member of the ALD subfamily, which is involved in peroxisomal import of fatty acids and/or fatty acyl-CoAs in the organelle. All known peroxisomal ABC transporters are half transporters which require a partner half transporter molecule to form a functional homodimeric or heterodimeric transporter. This peroxisomal membrane protein is likely involved in the peroxisomal transport or catabolism of very long chain fatty acids. Defects in this gene have been identified as the underlying cause of adrenoleukodystrophy; an X-chromosome recessively inherited demyelinating disorder of the nervous system (Kawaguchi and Morita, 2016).

2.7.3. Analysis of tissue-specific gene expression profiles *via* TiGER and HPA databases

The 10 genes (List 2) were subjected to cross cancer tissue specific analyses where normal and cancerous breast tissue gene expression were compared. Genes were used individually in TiGER and HPA as mentioned above in Sections 2.6.2.1 and 2.6.2.2 respectively.

The results in figures 2.1-2.3 displays a graphical output of three genes (ABCC5, ABCC10 and ABCC11) and their expression profiles calculated *via* an enrichment value in normal breast tissue and cancer tissue. The graphical data and expression profiles of the remaining seven genes can be found in Appendix A. with results provided for only three genes ABCC5, ABCC10 and ABCC11.

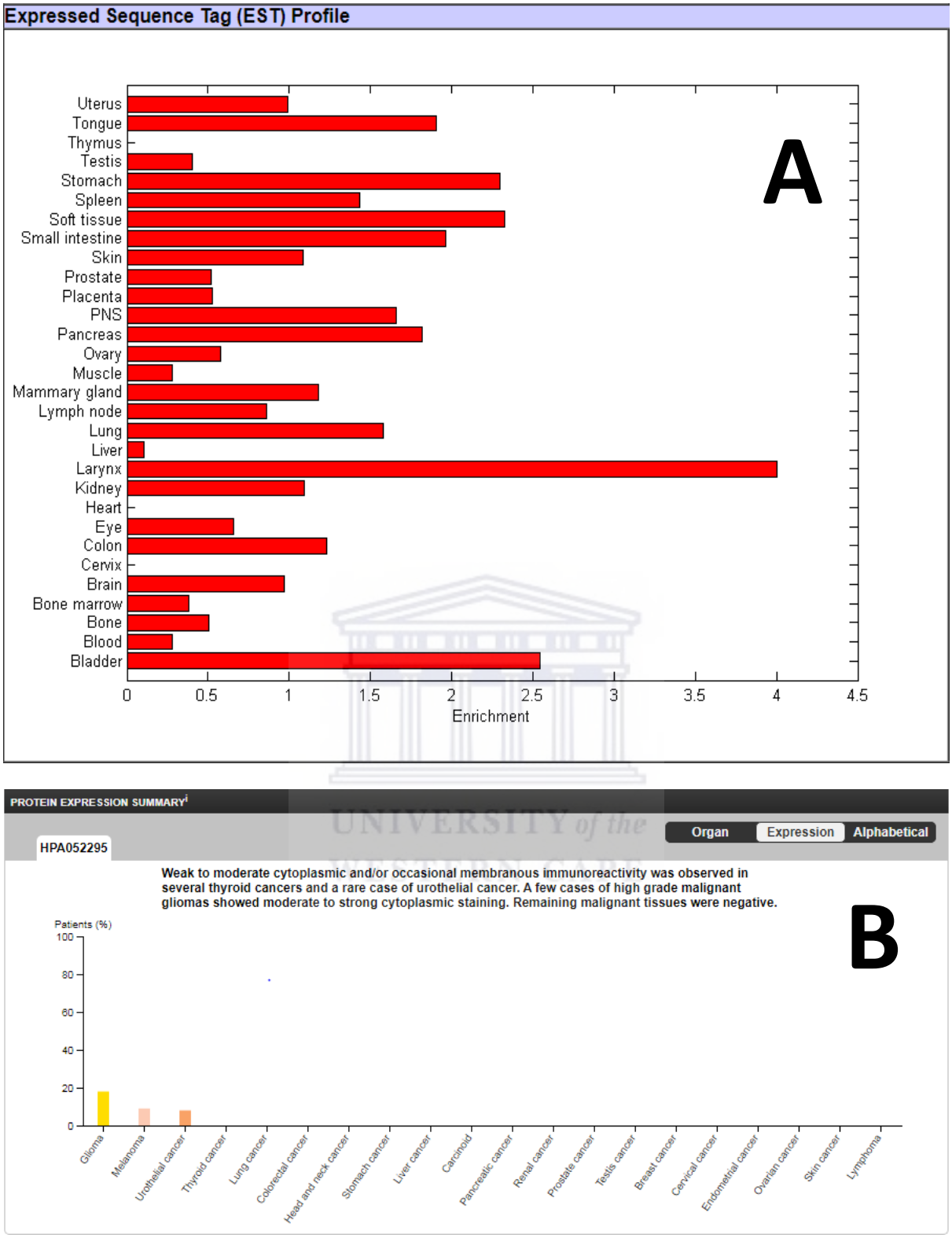

Figure 2.1: Expression profile for ABCC5 from TiGER (A) and (B) HPA.

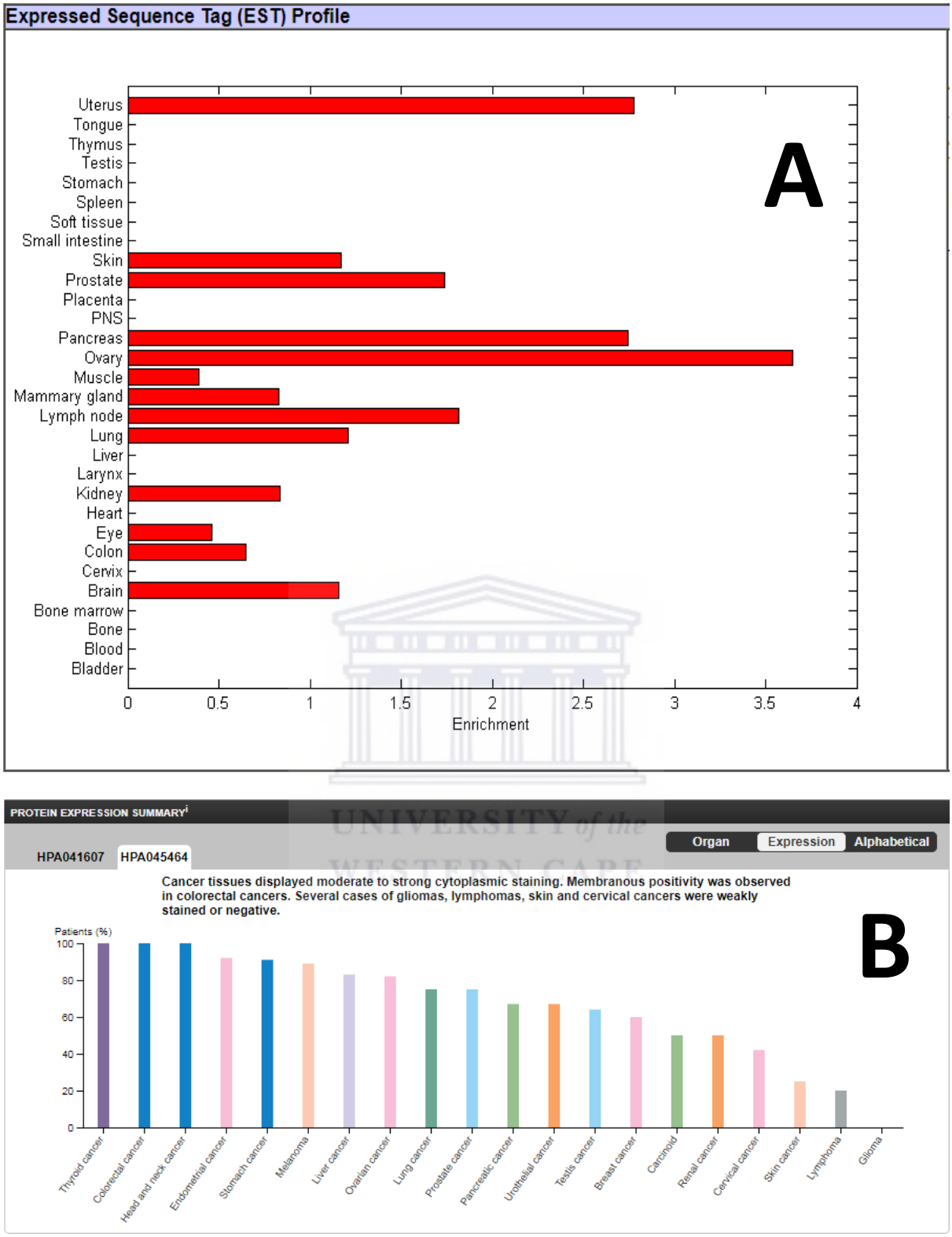

Figure 2.2: Expression profile for ABCC10 from (A) TiGER and (B) HPA.

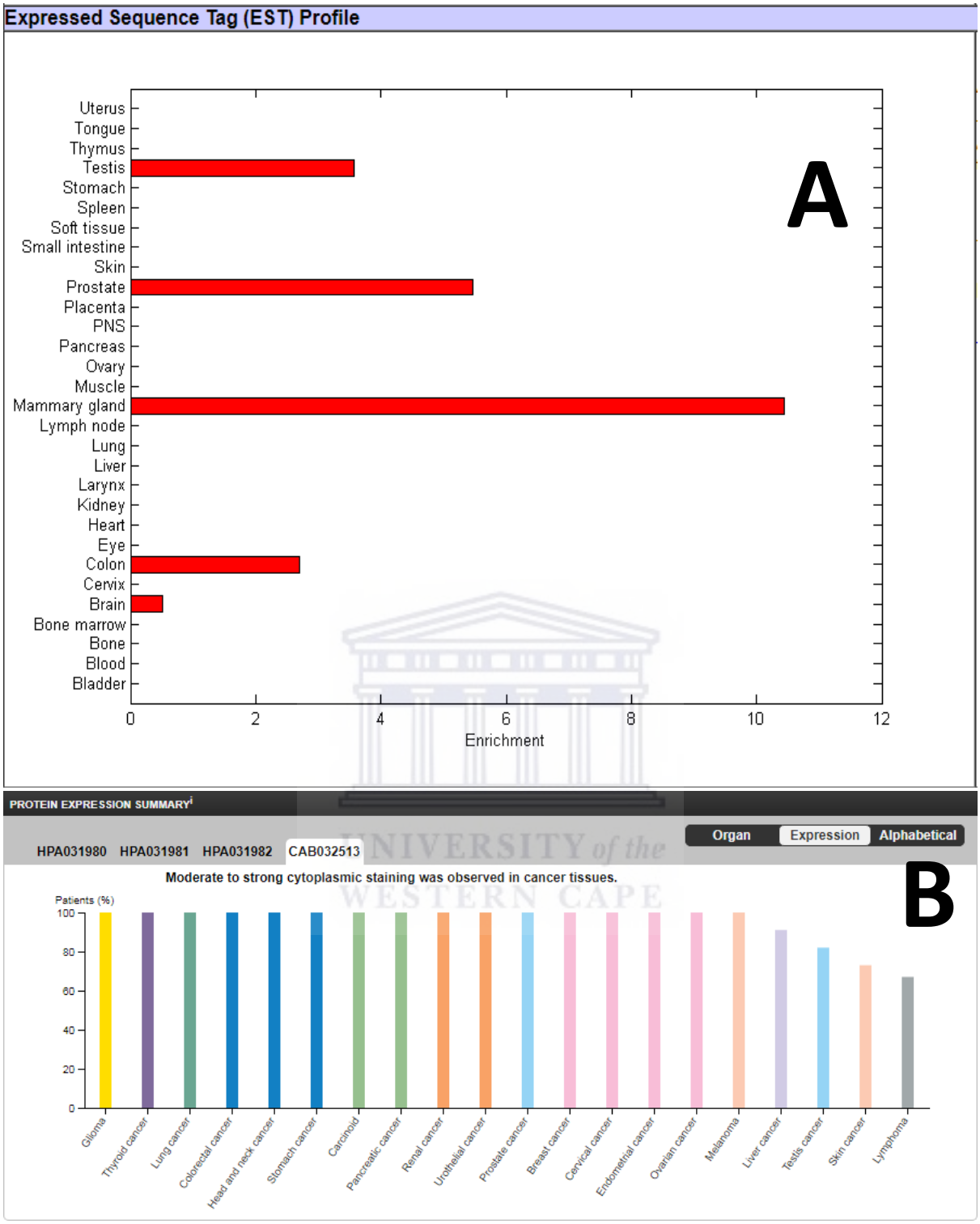

Figure 2.3: Expression profile for ABCC11 from TiGER (A) and (B) HPA.

Figure 2.1A shows the expression profile of the ABCC5 based on the results obtained from TiGER which suggest that this gene is preferentially expressed in the larynx followed by the bladder with an expression enrichment value higher than 4.0 and 2.5 respectively. Interestingly, ABCC5 was expressed in breast tissue (A) but was not expressed on breast

cancer (Figure 2.1B). However, the lack of expression data for ABCC5 in the HPA database could be ascribed to lack of experimental validation data in breast cancer, or it could suggest that gene expression in cancerous tissues are not well documented in the respective databases. The result observed by HPA is in contrast to what was observed by Park *et al.*, (2006), which demonstrated a high level of expression of ABCC5 in BC patients.

Figure 2.2 shows the expression profile of the ABCC10 based on the results obtained from TiGER (A) and HPA (B). The results from TiGER shows that ABCC10 is preferentially expressed in the ovaries, uterus and pancreas with an expression enrichment value greater than 2.5. The expression data obtained from HPA (B) shows that ABCC10 is overexpressed in breast cancer tissue. The upregulation of ABCC10 expression in breast cancer described here, is in support of a previous study by Park *et al.*, (2006), who demonstrated that several ABC transporters (ABCC5, ABCA12, ABCA1, ABCC13, ABCB6 and ABCC11), including, ABCC10, were significantly unregulated in the breast cancer patients with residual disease.

Furthermore, a study by Zhang *et al.* (2015), showed that several compounds were able to interact with ABCC10 and that ABCC10 can confer resistance to these compounds, both at the tumour cell level and by decreasing their oral bioavailability. Supporting a further role for ABCC10 in control of the response of tumours to the administration of therapeutics, Balaji *et al.* (2016), showed that overexpression of ABCC10 *in vitro* confers resistance to an unusually wide range of clinically valuable drugs, including taxanes, vinca alkaloids, and Tamoxifen.

Another study by Zhao *et al.* (2018), showed that ABCC10 expression level is elevated in non-small cell lung cancer (NSCLC) in relation to normal lung, with ABCC10 expression in adenocarcinoma correlated with tumour grade and stage providing further evidence for the role of this gene in tumourigenesis in both onset and progression. Thus, these studies correspond with and support the results obtained from this study with this gene being central as well in the network generated by STRING (Figure 2.4).

Figure 2.3 shows the expression profile of the ABCC11 based on the results obtained from TiGER results (A) and HPA (B). The results from TiGER shows that ABCC11 is preferentially expressed in the breast and prostate with expression enrichment values higher than 10.0 and 5 respectively. The expression data obtained from HPA result (B)

shows an over expression of the ABCC11 in breast cancer tissue. A previous study by Ota *et al.* (2010), showed that ABCC11 is potentially involved in conferring drug resistance to breast cancer and highly expressed in breast tumours in particular, in invasive ductal adenocarcinomas (Ota *et al*., 2010).

Its expression is reportedly regulated by ER-β and induced by 5-fluorouracil (5-FU). In addition, it has been reported that ABCC11 is directly involved in 5-FU resistance by the efflux transport of the active metabolite 5-fluoro-2'-deoxyuridine 5'-monophosphate (FdUMP) (Ishikawa *et al*., 2013; Sosonkina *et al*., 2011). 5-FU is a medication used to treat cancer and was patented in 1956 and came into medical use in 1962. 5-FU has been given systemically for anal, breast, colorectal, oesophageal, stomach, pancreatic and skin cancers especially head and neck cancers (Ishikawa *et al*., 2013; Sosonkina *et al*., 2011).

2.7.4. Gene-gene interaction networks using STRING database

Gene-gene interactions were analyzed for the shortlisted and prioritized genes of list 2 using the STRING Version 10 database as described in Section 2.6.2.3. Using the 'Multiple proteins' option, the gene IDs for the prioritized genes were used as input and searched for in '*Homo sapiens'* as the organism of choice (Sun *et al.*, 2005).

Figure 2.4 shows the gene-gene interaction between all prioritized genes that show up-regulation in breast cancer post-treatment. STRING, displays a score of confidence for each association between proteins (low confidence: scores <0.4; medium: 0.4 to 0.7; high: > 0.7), the higher the score the more evidence for the association recorded by STRING (Franceschini *et al*., 2013). In addition, each color line represents the evidence on which the interactions were made.

All the prioritized genes interact with each other as seen in figure 2.4. Gene's network interaction shows that most of these genes are sharing protein homology, co-expression and their interactions were identified through text mining as well (Section 2.7.2).

Protein-protein interaction networks or functional association networks are crucial in understanding the cellular machinery in organisms (Barabási and Oltvai, 2004), and can fulfill other practical purposes such as filtering and assessing high-throughput functional genomics data, and providing intuitive visual scaffolds for annotating the structural, functional and evolutionary properties of proteins.

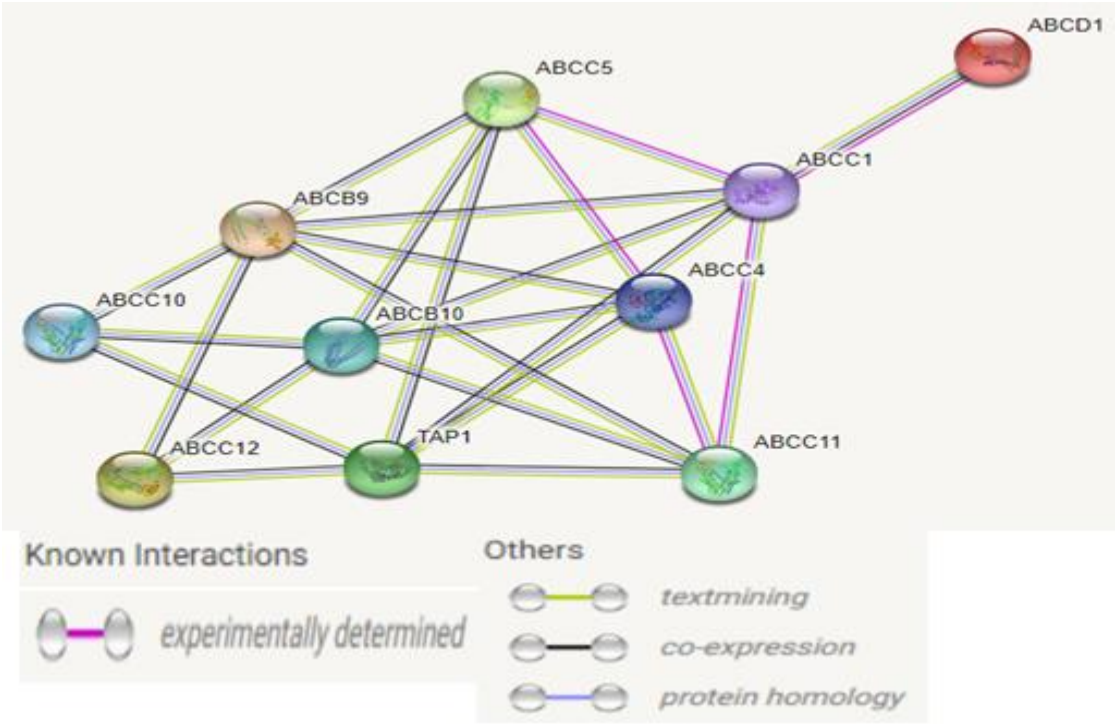

Figure 2.4: Gene-gene interaction network data representation generated by STRING.

Protein association networks have also proven surprisingly useful for the elucidation of disease genes, both for Mendelian and for complex diseases (Wang *et al*., 2011; Yamada *et al*., 2013). With the aim of this study being to identify ABC transporter genes involved in Tamoxifen resistance and in breast cancer post-treatment. By elucidating the pathways these genes are involved in and their interaction with other proteins within these pathways, a targeted therapeutic approach can be used to overcome the resistant mechanism by ABC drug transporter genes. This will lead to a better prognostic outcome for patients on this particular chemotherapeutic regiment.

Certain genes within the network have more than one association with a combined score close to one. The more combined associations between genes giving a score close to 1, the more evidence for the interaction (Franceschini *et al*., 2013). The more interactions a protein has in a network, the more important that protein is to the network as many other proteins rely on it for functioning (Franceschini *et al*., 2013).

For instant, ABCC4 is a member of the MRP family, acts as a regulator of intracellular cyclic nucleotide levels and as a mediator of cAMP-dependent signal transduction to the nucleus. ABCB10/M-ABC2/MTABC2 protein is a member of the MDR subfamily.

Members of the MDR subfamily are involved in MDR. In addition to that, ABCB10 is found in the inner membrane of mitochondria. In mammals ABCB10 is essential for erythropoiesis, and for protection of mitochondria against oxidative stress. Therefore, ABCC4 and ABCB10 genes seem to be important proteins within the interaction network as these genes are central in the produced network of interactions with three interactions identified between the prioritized genes.

Interestingly, the gene ABCD1 only interacts with one gene (ABCC1) in the interaction network with four interactions identified between the two genes. However, ABCD1 encodes a transporter protein of 745 amino acids, referred to as the adrenoleukodystrophy protein (ALDP). ALDP is an ATP-binding transport protein involved in the peroxisomal transport or catabolism of very long chain fatty acids which makes it a unique gene based on its distinguish location in X-chromosome and its function from the rest of the ABC subfamily (Kawaguchi and Morita, 2016).

2.7.5. Intersection analysis of Venn diagram of ABC transporter genes/Tamoxifen *via* CTD

Venn diagrams were generated by using the CTD online tool, representing the intersection between the prioritized ABC transporter genes and Tamoxifen. The intersecting ABC prioritized genes and Tamoxifen are shown below in Figure 2.5.

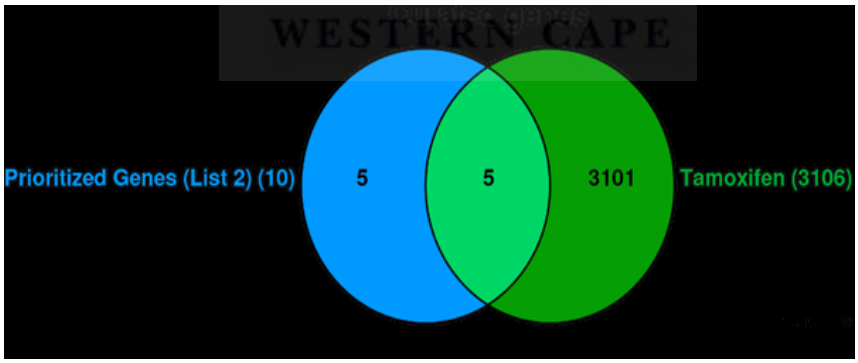

Figure 2.5: Venn diagram of the prioritized genes (List 2) interacting with Tamoxifen.

Figure 2.5 shows the intersectional relation between five out of 10 ABC transporter genes and Tamoxifen. The genes interacting within the intersection Venn diagram were ABCB2, ABCC1, ABCC4, ABCC11, and ABCD1. Several lines of evidence were presented for

ABCC11 as seen in the section above. Using a different *in silico* tool, the genes relevance for further investigation is very apparent through the results obtained.

2.8. Conclusion and summary

ABC proteins constitute a large family of active transporters through extracellular and intracellular membranes. The ABC transporters are a family of transporter proteins that are responsible for drug resistance and a low bioavailability of drugs by pumping a variety of drugs out of cells at the expense of adenosine tri phosphate hydrolysis. Drug resistance remains one of the primary causes of suboptimal outcomes in cancer therapy.

In this chapter, 10 ABC transporter genes as therapeutic biomarkers for breast cancer were identified.

A combination of biological text mining, data mining, and *in silico* gene enrichment techniques, proved to be effective in categorizing, classifying, and prioritizing the ABC transporter genes and linking them to breast cancer.

Tissue specificity expression analysis in TiGER revealed that ABCC11 is preferentially expressed in breast tissue. Results from HPA database showed that the ABC transporter genes were overexpressed in breast cancer tissue when compared to the normal tissue. The analysis also showed that ABCB9, ABCB9, ABCB10, and ABCC12 were all expressed in breast cancer tissue. However, the ABCC1, ABCC4 and ABCD1 genes obtained no results for gene expression in breast cancer tissue on the HPA database. ABCC12 gene obtained no result when was used individually with its Gene ID in the search option to show expression in breast tissue in the TiGER database. However, the absence of gene expression may be explained perhaps due to the fact that TiGER database either have partial information about the gene or may not contain enough experimental information to confirm the expression of this gene.

This development of computational methods for tissue-specific combinational gene regulation, based on transcription factor binding sites, CRMs and ESTs enables the platform to perform a large-scale analysis of tissue-specific gene regulation in human tissues. These comparisons will enhance understanding about the function of these genes and proteins and the molecular basis of differential susceptibility.

An analysis of STRING results indicated that the genes are involved in interaction networks which corresponds with the study that has been accomplished by Hlaváč *et al.*,

(2013), demonstrating that several ABC transporter genes were up-regulated in the breast cancer post treatment. Proteins/genes interact with one another not only by physical binding, but also indirectly by sharing an intermediary substrate in a metabolic pathway, by regulating the transcription of each other either through feedback or co-expression, or by participating in larger multi-protein assemblies.

In breast cancer patients, several ABC transporters, including ABCC5, ABCA12, ABCA1, ABCC13, ABCB6 and ABCC11, were significantly unregulated in the patients with residual disease. However, ABCC5 was found to be functionally involved in the formation of breast cancer bone metastases in two independent cell-based models. Therefore, these results support an important functional role for ABCC5 in promoting the colonization and growth of breast cancer cells specifically within the bone.

One recent study showed that ABCC10 mRNA is up-regulated in breast carcinoma and that its expression correlates with ER status (Hlaváč *et al*., 2013). Greater significance was observed in the ABCC10 expression levels of ER-positive lines compared with ER-negative lines. Also the study analysed ABCC10 mRNA level in two breast cancer data sets from Gene Expression Omnibus. ABCC10 expression was significantly higher in tumour than in normal lobular breast tissue in contrast to ABCB1 and ABCG2 expression. CTD database result centralizes the data that is core to toxicogenomics by integration and manual review of diverse molecular, reference, and chemical data. One of the primary goals of CTD is to advance the understanding of the effects of chemicals on human health on the genetic level, a field called toxicogenomics.

Based on the results from this chapter, it was concluded that the text and data mining, gene expression tissues-specificity analyses supported by the protein-protein network interaction as well as the comparative toxicogenomics could serve as a panel to identify biomarkers for diagnosis, prognosis or predictive markers for breast cancer management.

To validate these genes as biomarkers and therapeutic targets for successful treatment, their prognostic and predictive values were analyzed using different bioinformatics tools (See Chapter 3).

CHAPTER THREE: Survival, prognostic, and predictive analysis of the prioritized ABC transporter genes for breast cancer management using *in silico* approaches

3.1. Introduction

Clinical application and discovery of new biomarkers have been playing a significant role in reforming life science research and industry, in this manner intensely influencing the detection and treatment of many diseases, particularly in cancer (Bhatt *et al.*, 2010; Nalejska *et al.*, 2014). The National Cancer Institute defines biomarkers as any biological molecule that is isolated from body fluids such as urine, blood or tissue that is a sign of a normal or abnormal process in the body or disease condition. They are objective indications of a medical state observed from outside the patient, which can be measured accurately and is reproducible (Kyle strimbu and Jorge A. Tavel, 2010).

Biomarkers in the biomedical field and clinical application have fundamentally changed the medical field because they are clinically characterized as effective non invasive indicator molecules (Chatterjee and Zetter, 2005; Prassas, Chrystoja, Makawita, and Diamandis, 2012). Therefore, the discovery of cancer biomarkers has become an important part in biomedical research in favor of utilization within the medical practice (Bhatt *et al*., 2010). For instance, in the case of cancer, biomarkers will be useful for accurate evaluation of a disease stage and can be used for predicting several outcomes during the course of disease including: early detection, outcome prediction and detection of disease recurrence (Chatterjee and Zetter 2005).

The ideal cancer biomarker should be easily accessible, highly specific, and sensitive enough to detect tumours at its different stages. 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 (Bhatt *et al*., 2010).

Identification of biomarkers requires a study of the involved 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, prognostic, and therapeutic biomarkers (Prassas *et al*., 2012; Soleas *et al*., 1997).

Identifying and validating diagnostic, prognostic, predictive, and therapeutic biomarkers as well as the stage that its being identified at, whether genetic or metabolic, will have a huge impact on the patient's outcomes as they will allow early detection of tumours and also guide the choice of a targeted therapy based on specific molecular features of the cancer (Figure 3.1) (Chatterjee and Zetter, 2005; Nalejska *et al.*, 2014).

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 and Diamandis, 2008). A diagnostic biomarker as defined by Carlomagno *et al.,* (2017), allows for early and non invasive detection of cancer including secondary prevention of cancer.

Despite current research advancements, a huge gap in the presence of such diagnostic biomarkers, specifically for early disease detection, such as cancer, exists. While conventional diagnostic biomarkers largely involve histopathological variables, recent methodologies include the use of molecular biomarkers (Carlomagno *et al*., 2017). 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 (Nalejska *et al*., 2014). Recent studies have focused on the use of circulating miRNAs as diagnostic signatures in disease identification, including breast cancer (Blenkiron *et al*., 2007; Iorio *et al*., 2005).

3.1.2. Predictive biomarkers

A predictive biomarker eases the monitoring of patients during treatment to measure the response of the patient to the specific treatment. Additional considerations may apply when evaluating the clinical utility of a predictive biomarker in selecting between two therapy options (Burke 2016).

The emergence of innovative technologies in Bioinformatics, Genomics, Proteomics, Metabolomics and imaging, allows researchers to facilitate inclusive analysis of cancer cells. These approaches have already established its power to distinguish cancer cells from

normal cells and to identify specific genetic elements involved in cancer (Kontos *et al.*, 2015).

Predictive biomarkers represent a treatment response prospective and guide clinicians in taking therapeutic decisions (Bhatt *et al.*, 2010; Nalejska *et al.*, 2014). Commonly, somatic mutations in genes like epidermal growth factor receptor (EGFR), KRAS, BRAF, PDGFRA, KIT, HER-2, BCRABL, 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 EGFR, using Cetuximab or Panitumumab in colorectal carcinogenesis (Walther *et al.*, 2009; Nalejska *et al.*, 2014).

Similarly, gene polymorphism in CYP2D6*10/*10 (and CYP2D6*5/*10) and low concentrations of Endoxifen and 4-hydroxy-Tamoxifen, the active metabolites of Tamoxifen has been correlated with weak drug response in breast cancer patients (Wang *et al*., 2011; Yu *et al*., 2006).

3.1.3. Prognostic biomarkers

Prognostic biomarkers are characterized as biomarkers that are assigned to a specific tumour type by either, determining the occurring polymorphism, mutation or the change in DNA methylation or gene expression. Prognostic biomarkers should be able to 1) detect the presence of specific miRNA molecules or Circulating Tumor Cells (CTC) in the peripheral blood with the ability to assess treatment and disease outcomes (Nalejska *et al.*, 2014) and 2) be highly specific to detect cancer stages and disease remission prognosis (Carlomagno *et al*., 2017).

Furthermore, the prognostic biomarker is a clinical or biological characteristic that provides information on the likely course of the disease; it gives information about the outcome of the patient (Tabarestani *et al*., 2014a). However, prognostic biomarker concentration should directly implicate the disease stage and therapeutic outcomes of anti-cancer drugs for an individual.

Tumour classification, staging and sometimes grade is normally used to assess prognosis. Nonetheless, staging could incorporate other parameters that improve prognosis bringing about additional cost (Nalejska *et al.*, 2014; Duffy *et al.*, 2017).

Biomarker expression often replaces or complements tumour classification, stage and grade when biologically targeted therapeutics is under consideration, but addition of markers could similarly fragment the staging process, thereby limiting its utility (Ludwig and Weinstein, 2005).

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 breast cancer for the assessment of treatment outcomes, cancer proliferation and invasiveness in breast cancer using microarray, qRT – PCR and *in silico* tools. Additionally, the presence of specific miRNAs has been correlated with several cancers (Badve *et al*., 2011; Mills *et al*., 2011; Van De Vijver *et al*., 2009).

DNA methylation has also been reported to act as a potential prognostic marker as hypermethylation patterns of gene encoding heat-shock proteins 27 and heat-shock protein 1 could serve as a good prognostic biomarker (Lorincz *et al*., 2013). In-extension, 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 (Easton *et al*., 2007). In a study conducted by Zhang *et al.*, (2012), the presence of five or more CTCs was correlated with overall survival in prostate cancer patients.

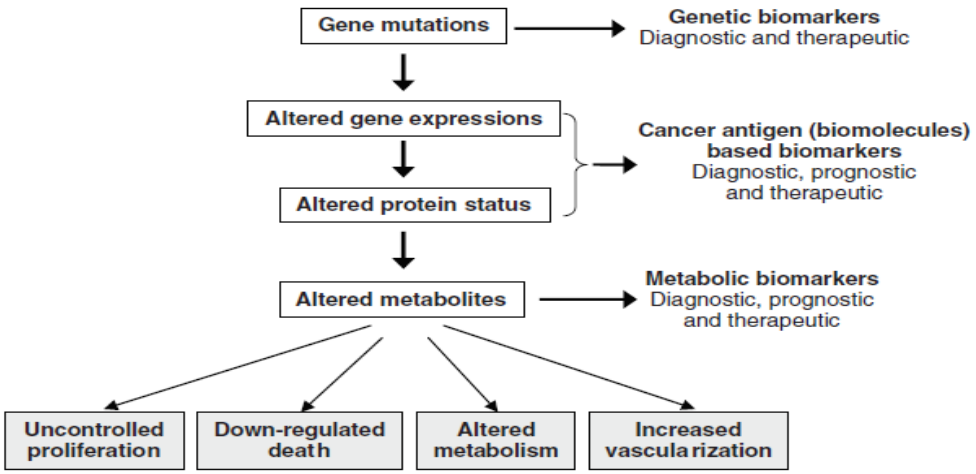

Figure 3.1: The process of carcinogenesis, showing opportunities of identifying biomarkers (Adopted from Bhatt *et al.*, 2010).

3.1.4. Breast cancer biomarkers

Conventional breast cancer 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 Tumour Marker Utility Grading System. These markers include the ER and human epidermal growth factor receptor 2 (HER2) (Easton *et al*., 2007; Weigel and Dowsett, 2010).

ER expression plays a crucial role as a predictive biomarker as the absence or presence of the ER provides information about endocrine treatment outcomes (Hamed *et al*., 2019). Furthermore, the expression of ER strongly influences the expression of PR. Several studies on adjuvant trials reported a strong prognostic value but mild predictive significance for PR expression (Dowsett *et al*., 2006; Weigel and Dowsett, 2010).

Oncogene HER-2 was identified as a prognostic biomarker in breast cancer, due to the augmented expression levels detected in breast cancer patients. Such patients are highly susceptible to disease recurrence and have a shorter overall survival (Chatterjee and Zetter, 2005). Antigen KI-67 (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 maximally at M phase (Colomer *et al*., 2018). Ki67 has been correlated with other breast cancer biomarkers like cancer invasiveness, tumour grade and ER (Colomer *et al*., 2018).

Several research findings implicate Ki67 as a strong predictor of luminal A and B subtypes, chemotherapy response, disease recurrence and overall survival in breast cancer patients (Duffy *et al*., 2017; Weigel and Dowsett, 2010).

3.2. Cancer bioinformatics tools in survival, 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 (Aguirre-Gamboa *et al*., 2013; Bhatt *et al*.,

2010). Several bioinformatics tools have been developed to assess the predictive, prognostic and therapeutic utility of candidate biomarkers such as: Kaplan Meier plot, Recurrence Online, breast cancer GeneExMiner, Gene Expression-Based Outcome for Breast Cancer Online (GOBO) and PrognoScan (Aguirre-Gamboa *et al*., 2013).

In the present study, SurvExpress and PROGGENE databases were used to study the survival and prognostic gene expression analysis and they are described in details below.

3.2.1. SurvExpress database

SurvExpress is a sophisticated gene expression database and bioinformatics tool equipped to generate survival analysis and risk assessment of cancer datasets (Györffy *et al*., 2010). The tool can be accessed at http://bioinformatica.mty.itesm.mx/SurvExpress and has the ability to validate survival and prognostic cancer biomarkers (Aguirre-Gamboa *et al*., 2013). The gene expression database comprises 20,000 samples from 20 different types of cancers along with 130 datasets. The candidate genes are used as input in the web interface and the datasets used to evaluate gene expression, are selected from 140 available datasets (Aguirre-Gamboa *et al*., 2013).

Györffy *et al.* (2010) described SurvExpress as an online biomarker validation tool. They indicate 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 (Aguirre-Gamboa *et al*., 2013). The significance level value using a t-test or f-test through box plots across risk groups and the relation between risk groups and prognostic indices are indicated with the result outputs.

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).

3.2.2. PROGGENE database

PROGGGENE database is a web application tool that can be used to study prognostic implications of genes in various cancers. PROGGENE database is a publicly available online database that is accessible at http://genomics.jefferson.edu/proggene.

PROGGENE database is considered as a hypothesis generation tool for researchers to identify potential prognostic biomarkers and studying prognostic implications of mRNA

biomarkers in a variety of cancers. The generated prognostic plot is being created by using R library "survival" package with the aid of the compiled data from public repositories such as Gene Expression Omnibus, European Bioinformatics Institute, Array Express, and The Cancer Genome Atlas (Goswami and Nakshatri, 2013).

PROGGENE database as a bioinformatics tool uses the data pertaining to specific patient in its population which allows researchers to study biomarkers in the specific patient populations and the outcome data provides the most comprehensive resource available for survival analysis to date (Goswami and Nakshatri, 2013).

3.3. Aim and objective

The aim of this chapter was to evaluate the survival, predictive and prognostic value of the prioritized ABC transporter genes in breast cancer management.

Objectives

1. Prognostic analysis of candidate genes using datasets from SurvExpress database.
2. Over Survival analysis of candidate genes for different breast cancer types using PROGGENE database.
3. Predictive of metastasis validation of candidate genes using datasets from PROGGENE database.

3.4. Methods and materials

3.4.1. Validating the ABC candidate genes for their prognostic impact using datasets from the SurvExpress database

SurvExpress is a cultured and refined database for performing biomarker validation for cancer gene expression analysis across different cancer tissues. SurvExpress accessed at http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp was used to evaluate the prognostic value(s) of the 10 ABC transporter genes in breast cancer patients using the following steps: (i) ABC gene identifiers was used as input into the "query" option; (ii) SurvExpress incorporates several databases and breast cancer recurrence data, nine datasets from seven authors was selected, as it houses the largest number of samples totaling 1574, (iii) under the duplication option two settings were selected, (a) "show all" and (b) data "original.

SurvExpress analysis was applied with censored data set to survival-days. In the following opened web page, biomarker analysis censored was set up as "Tamoxifen therapy post-surgery" and stratification was selected as "Tamoxifen treatment post-surgery" then the "Go" button was pressed to complete the analysis. The results were displayed as graphical outputs of a gene and survival calculated using a statistical significant *P* value ($P < 0.05$, 0.01 and 0.001).

The effect of low or high expression of a biomarker was graphically represented and the results saved in PDF format. Data of only three representative genes were displayed in the results section whereas the remaining genes were documented in Appendix B.

3.4.2. Validating the ABC candidate genes for their over survival and metastasis free survival impact using datasets from PROGGENE database

PROGGENE database is a web-based application tool used to study prognostic, survival and predictive implications of genes in various cancers. PROGGENE accessible at http://genomics.jefferson.edu/proggene/ was used to perform the analysis for the ABC transporter genes by submitting each gene individually into the input gene(s) section option. From the list of different cancers provided, "breast cancer" was selected.

Survival measurement was selected as "death" to study the over survival analysis and "metastasis" was selected to study the metastasis free survival expression analysis and the bifurcate gene expression was selected as "median". Based on the different filter selections, datasets were specified to create the plots as a Kaplan Meier plot. Result figures displayed a

graphical output of a gene and survival calculated using a statistical significant *P* value ($P < 0.05$, 0.01 and 0.001).

The effect of low or high expression of the biomarker was graphically represented and the output saved in PDF format. Data of three representative genes were displayed in the results section with the data for the remaining genes displayed in Appendix B.

3.5. Results and discussions

The prioritized ABC transporter genes were analyzed *via* the SurvExpress and PROGGENE databases which are regarded as breast cancer online platforms utilizing the available clinical datasets as explained in Section 3.4.1 and 3.4.2.

3.5.1. Gene expression based survival analysis on the ABCC4 gene using different databases

Gene expression based prognostic, over survival and metastasis free survival analysis were studied in the above-mentioned databases and the results shown below in figure 3.2 displaying a graphical output of the combination results from two databases (SurvExpress and PROGGENE) and their survival expression profiles calculated *via* a statistical significant *P* value in breast cancer tissue. The graphical data profiles of the remaining genes can be found in Appendix B. Except for the ABCB9, ABCB10, ABCC11 genes obtained no results when their genes symbols were used individually in the SurvExpress and PROGGENE databases.

3.5.1.1. Prognostic analysis *via* SurvExpress on ABCC4

Result obtained from SurvExpress database (Figure 3.2A) showed the examination of prognostic and gene expression analysis for ABCC4 in breast cancer patients. The result showed a continuous decrease in patient numbers and creates a noticeable gap in survival rates for both high and low expression levels after 5 years. Within the first 5 years, the plot had a *P* value ≤ 0.01 thus making it a good examination of survival measure while being statistically significant.

A study done by Bagnoli *et al.* (2013), revealed preclinical proof supporting that the expression of ABCC1/MRP1 and ABCC4/MRP4 is up-regulated in ovarian carcinoma cells showing resistance to platinum drugs. The study revealed a relationship between MRP1 and disease grading, it was also noted that MRP4 showed an unfavorable progression of the disease outcome (Bagnoli *et al*., 2013). In another study by Murray *et al.,* (2017), on patients with neuroblastoma tumours, the up-regulated gene expression of ABCC1 and ABCC4 leads to poor clinical outcomes.

Therefore, the study supports the obtained results on ABCC4 that high level of this gene's expression affects recurrence free survival in patients negatively and thus lead to poor

prognosis. This supports the idea that ABCC4 would be a potential prognostic biomarker for breast cancer.

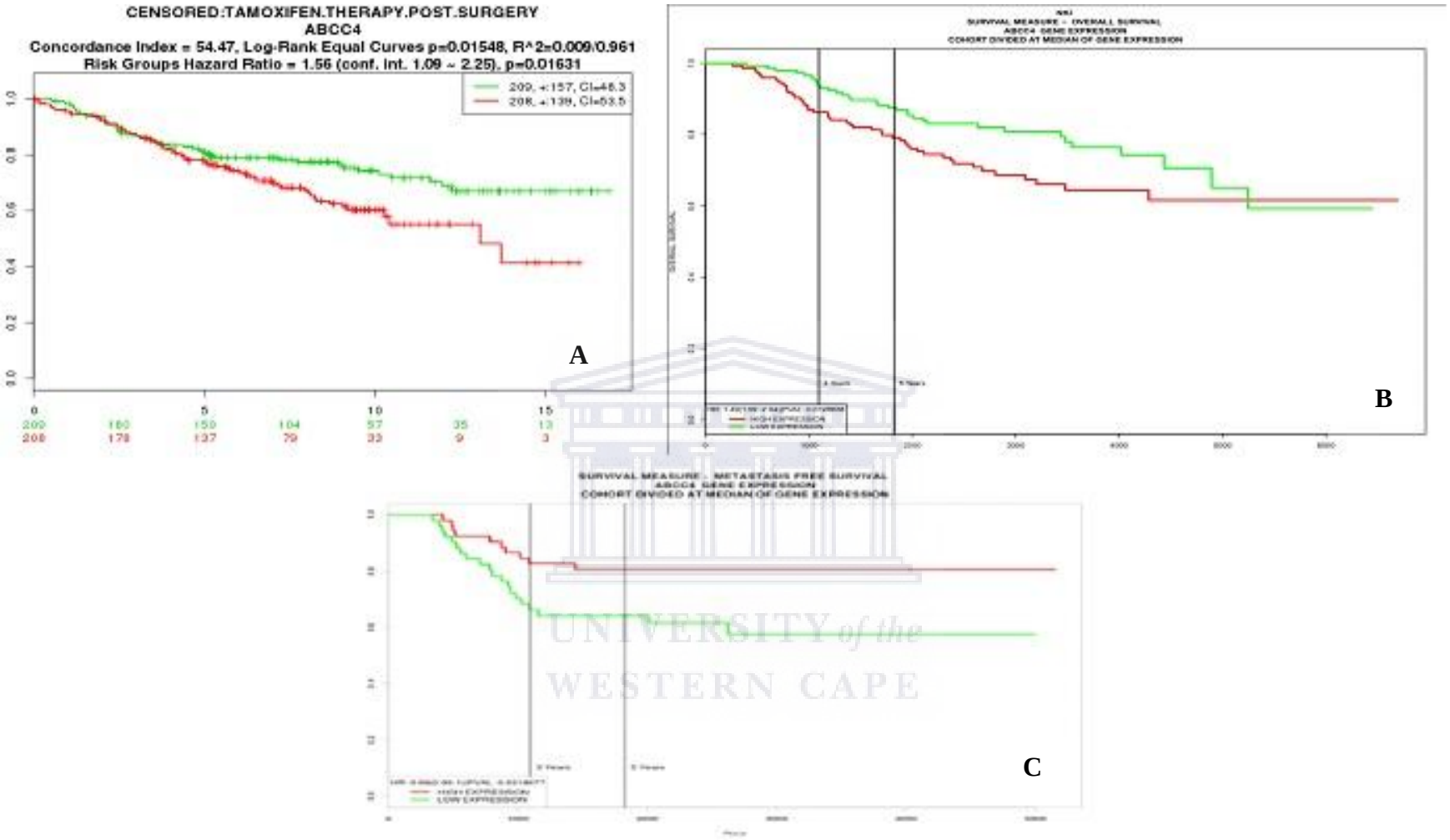

Figure 3.2: Survival analysis profile for the ABCC4 gene using SurvExpress (A) and PROGGENE (B and C) databases. The expression analysis captured in these plots shows the low (green) and high (red) gene expression in risk groups respectively using the log rank test (*P* value).

Enhanced expression of ABCC1 and ABCC4 in the cells dissected from metastatic tissues and metastatic lymph nodes as opposed to cells acquired from primary tumours, suggests a possible contribution of different ABC transporters in metastatic spread (Domenichini *et al*., 2019). Past examinations propose ABCC4 as a potential therapeutic target, as its knockdown suppresses multiplication of a scope of disease cell lines *in vitro* and xenograft models (Murray *et al*., 2017).

3.5.1.2. Over survival analysis *via* PROGGENE database on ABCC4

The plot obtained from PROGGENE database (Figure 3.2B) shows the mRNA gene expression level of ABCC4 gene through measuring the over survival for breast cancer patients. However, the result obtained for ABCC4 showed a decrease in survival rates for both high and low expression levels of this gene within the 1100 days creating a noticeable gap between the two graphs. The plot shows a P value ≤ 0.01 thus the over survival index value is considered highly significant and the survival measurement result makes it a good analysis.

A previous study has reported that, the failure of conventional and targeted cancer therapy can occur through an increased expression of efflux of chemotherapeutics, leading to reduced intracellular drug levels and consequently drug insensitivity, usually to multiple agents (Pérez-Herrero and Fernández-Medarde, 2015). Therefore, a well-documented reason of cancer cell MDR is through the increased expression of ABC family members, specifically, the ABC subfamily C which can export a variety of chemotherapeutics out of the cell (Begicevic and Falasca, 2017).

3.5.1.3. Metastasis free survival analysis *via* PROGGENE on ABCC4

The result obtained from PROGGENE database (Figure 3.2C) shows mRNA prognostic high expression level of ABCC4 gene through examining the metastasis free survival for breast cancer patients. The result obtained for ABCC4 showed a decrease in survival rates for both high and low expression levels of this gene within 3 years. The plot had a P value ≤ 0.01 thus the prognostic index value is considered highly significant and the survival measurement result, which makes it a good analysis.

Studies have recommended that ABCC4 is a valuable prognostic marker for neuroblastoma and is an androgen-regulated gene that is significant in prostate malignant growth (Zhang *et*

al., 2012). In this manner, the up regulation of ABCC4 in malignant growths proposes the likelihood that ABCC4 could assume a functional role in disease progression (Zhang *et al*., 2012).

Another study demonstrated that high levels of ABCC4 expression specifically in human pancreatic tumour tissues, may add to a high-chance of pancreatic malignant growth. In addition, ABCC4 advanced human pancreatic malignancy cell multiplication and state development *in vitro* by adjusting the cell cycle (Zhang *et al*., 2012).

Inhibition of ABCC4 is a potential approach for chemosensitisation in disease, justifying to the further advancement of small-molecule ABCC4 inhibitors (Murray *et al*., 2017).

It has clearly been demonstrated from this study and others that ABCC4 differential expression is associated with various cancer progression, metastasis as well as chemo-sensitivity thus it needs to be ascertained in which of the cancers the gene is most specifically expressed for it to be considered of prognostic value in breast cancer.

3.5.2. Gene expression based survival analysis on the ABCC5 gene using different databases

Gene expression based prognostic, over survival and metastasis free survival analysis were studied in the above-mentioned databases and the results in figures 3.3 display a graphical output of the combination results from two databases (SurvExpress and PROGGENE) and their survival expression profiles calculated *via* a statistical significant value in breast cancer tissue. The graphical data profiles of the remaining genes can be found in Appendix B. Except for the ABCB9, ABCB10, ABCC11 genes that obtained no results when their gene symbols were used individually in the SurvExpress and PROGGENE databases.

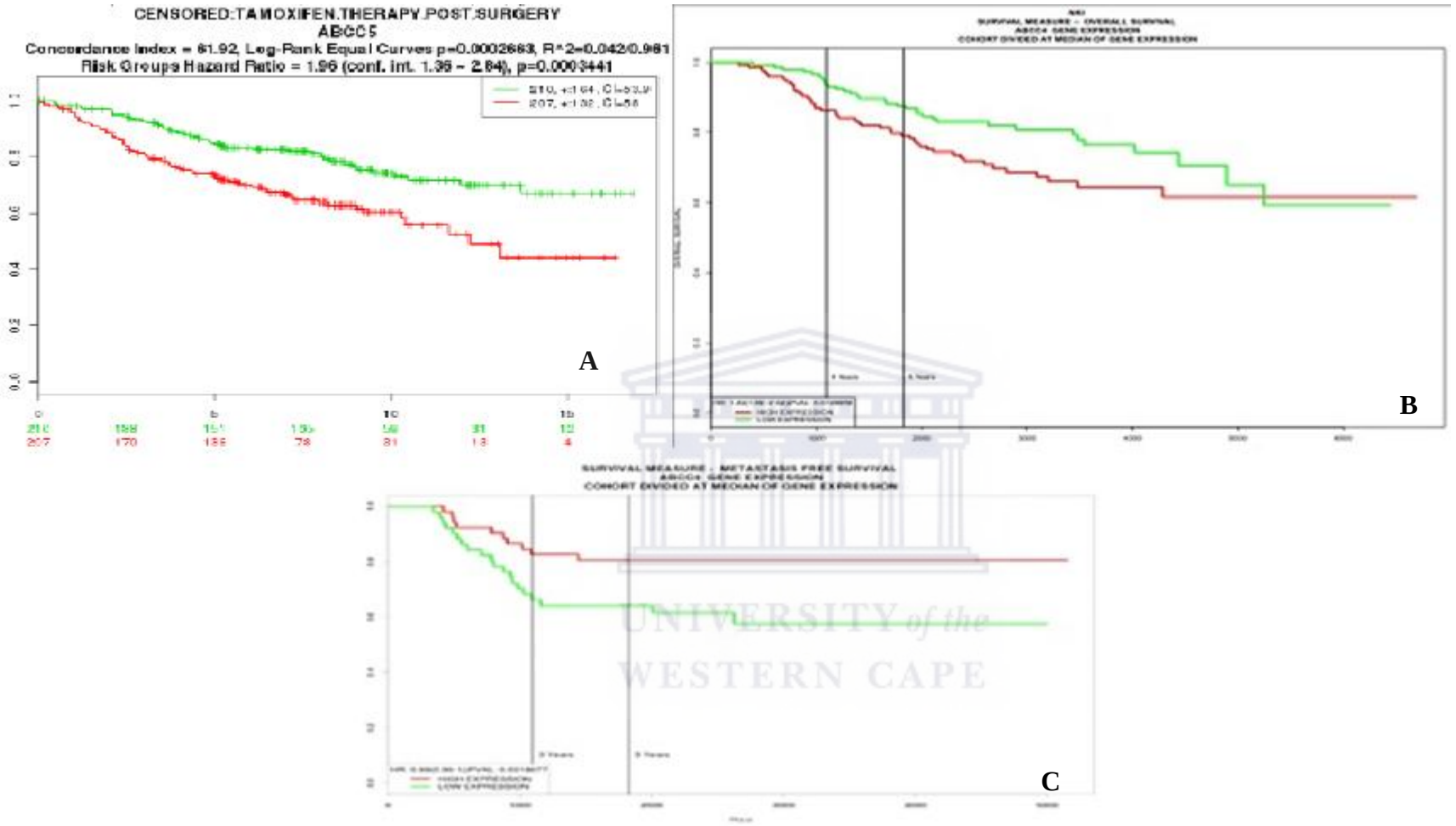

Figure 3.3: Gene Expression based Survival analysis profile for the ABCC5 gene using SurvExpress (A) and PROGGENE (B and C) databases. The expression analysis captured in these plots shows the low (green) and high (red) gene expression in risk groups respectively using the log rank test (*P* value).

3.5.2.1. Prognostic analysis *via* SurvExpress database on ABCC5

Result obtained from SurvExpress database (Figure 3.3A) shows the examination of the prognostic gene expression for ABCC5.

In breast cancer patients the result obtained shows a continuous decrease in the graphs obtained and created a noticeable gap in survival rates probabilities for both high and low expression levels of this gene, within the first 5 years. The plot (Figure 3.3A) shows a *P* value ≤ 0.001 therefore making it a good representation of survival measure while being statistically significant.

When ABCC5 is expressed at low levels the 10 years's survival rate is more than 70%. This supports the notion that ABCC5 would be a prognostic biomarker for breast cancer as high levels of its expression suggest a rapid decline in the survival rate. Therefore, over survival for breast cancer patients with high expression of the ABCC5 gene post Tamoxifen treatment correlates with poor clinical outcomes.

Results showed that ABCC5 over-expression correlated with an unfavorable outcome for prostate cancer patients. The prognostic value of ABCC5 was statistically significant in a multivariate analysis, which recommend ABCC5 over-expression might be a good molecular marker to predict prostate cancer prognosis (Zhang *et al*., 2018). Another study by Borel *et al*., 2012, showed the up-regulation of the ABCC5 in hepatocellular carcinoma and unfavorable outcome in liver cancer patients.

3.5.2.2. Over survival analysis *via* PROGGENE database on ABCC5

The result obtained from PROGGENE database (Figure 3.3B) shows mRNA gene expression profile of ABCC5 through examining the over survival for breast cancer patients. The result showed a decrease in survival rates for both high and low expression levels of this gene within 3 years creating a noticeable gap between the expression levels The plot shows a *P* value ≤ 0.01 thus the over survival index value is considered to be highly significant as well as the survival measurement result, which makes it a good result for this particular gene.

A previous study has detailed that failure of regular and focused cancer treatment can occur through an expanded expression of certain proteins that leads to an efflux of chemotherapeutics, prompting decreased intracellular medication levels and thus decrease the efficacy of the drug for various medications (Lal *et al*., 2017). This supports the notion

that ABCC5 is a potential prognostic biomarker for breast cancer as high levels of its expression results in a rapid decline of over survival in breast cancer patients. An earlier study has shown that, the high expression level of ABCC5 showed metastasis free survival in breast cancer patients and poor prognosis following treatment (Eggen *et al*., 2012).

Subsequently, in past examinations, it has been shown that the clinical tumour's reaction to Neoadjuvant Chemotherapy (NAC), for example, Tamoxifen, indicated that it was related with changes in the expression of the MDR gene, for example, ABCC5, in breast cancer tissue after the use of chemotherapy. Moreover, it has been demonstrated that a favorable tumour reaction is observed, which is related to MDR gene down-regulation within breast tumours after NAC.

Interestingly, poor clinical reaction to NAC was observed to be related to the increased expression of certain MDR genes in tumour tissue after NAC and a poor prognosis. In this manner, decreased MDR gene expression in breast tumours after Neo-adjuvant chemotherapy is related with its efficacy, yet the molecular mechanism for MDR down-regulation is inadequately comprehended (Lal *et al*., 2017).

3.5.2.3. Metastasis free survival analysis *via* PROGGENE on ABCC5

The result obtained from PROGGENE (Figure 3.3C) shows high mRNA expression level of ABCC5 through examining the metastasis free survival for breast cancer patients. This graph indicates that high expression for ABCC5 correlates with poor prognosis and metastasis free survival. The outcomes were compared with a past study by Mourskaia *et al.*, (2012) showing that, ABCC5, an ATP-dependent transporter, was observed to be overexpressed in breast malignant growth and promotes osseous metastases for bone cancer spreading from the primary breast tumours.

In an investigation by Yoshida *et al.*, (2001), examination of ABCC5 mRNA levels in doxorubicin-resistant human lung malignancy cells SBC-3/ADM, AdR MCF-7 and K562/ADM demonstrated particularly high expression of ABCC5 transcripts with respect to their individual parental cell lines. Another study by Estudante *et al.*, (2016), further showed that ABCC5-transfected HEK cells had a two-fold higher resistance for doxorubicin than non-transfected cells. Taken together, these discoveries propose that ABCC5 expression and activity may add to fluctuation in doxorubicin disposition observed in malignant growths (Lal *et al*., 2017). Doxorubicin is commonly used to treat

solid cancers of the bladder, breast, stomach, lung, ovaries, thyroid and soft tissue sarcoma (Lovitt *et al*., 2018).

Later investigations have highlighted the potential significance of different transporter proteins in balancing the disposition of doxorubicin, including ABCB5, ABCC5 and RalA-binding protein 1 (Lal *et al*., 2017). ABCC5 has been shown to intervene in the ATP-dependent transport of a few anticancer medications, including doxorubicin, and has been accounted for conferring resistance for various chemotherapy agents, including methotrexate, and the thymidylate synthase inhibitor raltitrexed (Lal *et al*., 2017).

3.5.3. Gene Expression based survival analysis on the ABCC10 gene using different databases

Gene expression based prognostic, over survival and metastasis free survival analysis were studied in the above-mentioned databases and the results below in Figures 3.4 display a graphical output of the combination results from two databases (SurvExpress and PROGGENE) and their survival expression profiles calculated *via* a statistical significant value in breast cancer tissue. The graphical data profiles of the remaining genes can be found in Appendix B. Except for the ABCB9, ABCB10, ABCC11 genes obtained no results when their gene symbols were used individually in the SurvExpress and PROGGENE databases.

3.5.3.1. Prognostic analysis *via* SUREXPRESS on ABCC10

Result obtained from SurvExpress database (Figure 3.4A) shows the examination of prognostic gene expression for ABCC10 in breast cancer patients. The result obtained from SurvExpress database shows a continuous decrease in survival rates for both high and low expression levels starting from the first month of survival until after 10 years and within the first 5 years.

The plot showed a P value ≤ 0.01 thus making it a good visualization of survival measure while being statistically significant. After 5 years, for the group of patients who showed low expression of ABCC10, more than 80% survived, which suggest that high expression of this gene correlates to a poor prognostic outcome.

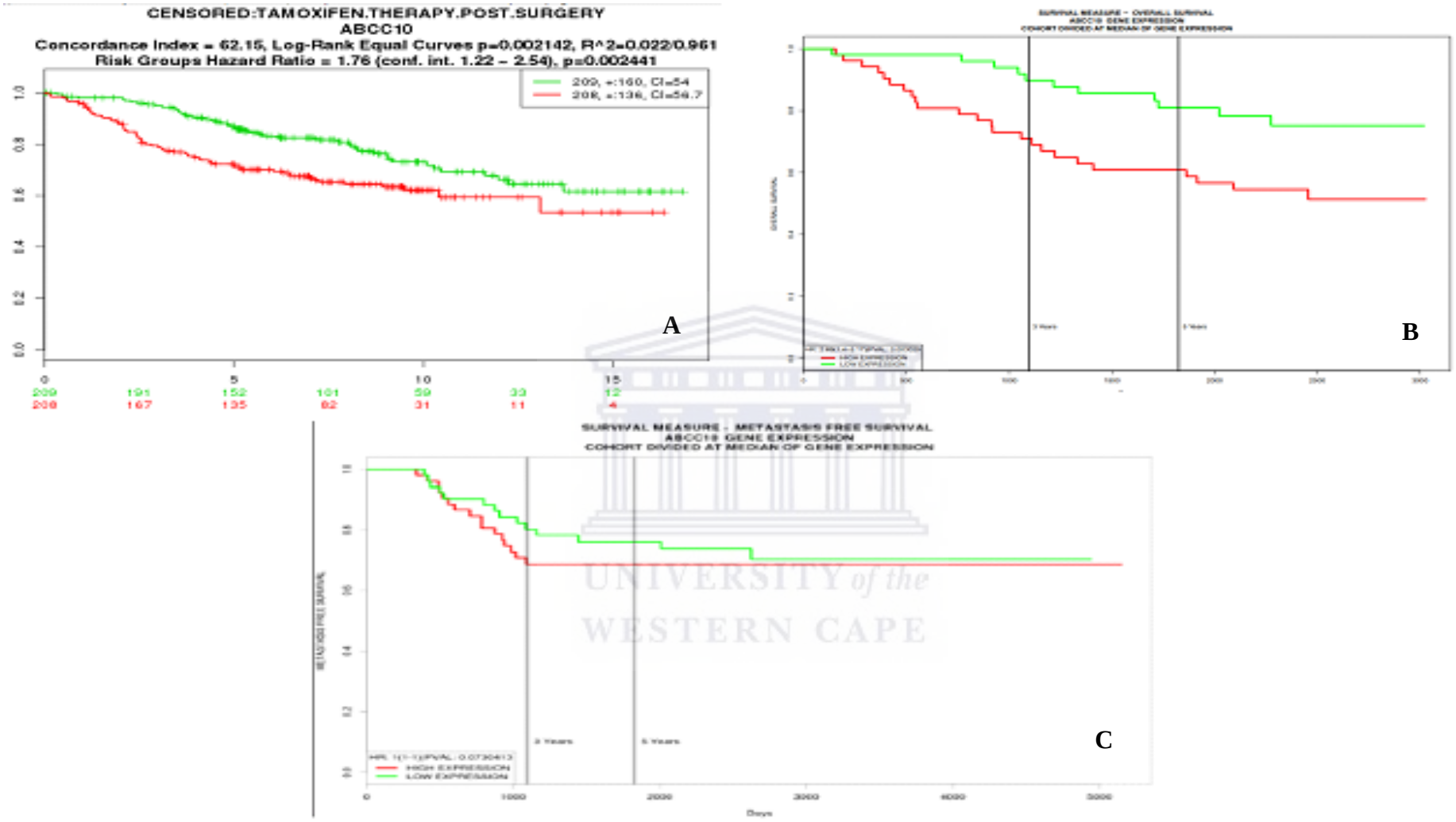

Figure 3.4: Gene Expression based Survival analysis profile for the ABCC10 gene using SurvExpress (A) and PROGGENE (B and C) databases. The expression analysis is captured in these plots shows the low (green) and high (red) gene expression in risk groups respectively using the log rank test (*P* value).

Knowledge of ABCC10 expression is significant for understanding the potential impacts of its efflux action and activity in normal and tumour tissues. ABCC10 transcript expression is widespread and broad, with the highest levels distinguished in the gonads, testis and spleen (Kruh *et al*., 2007).

ABCC10 transcript has also been detected in breast, lung, colon, ovarian, and pancreatic tumours, despite the fact that these investigations might have been constrained by a small sample size. In addition, ABCC10 transcripts have been identified in the HepG2 liver cancer cell line and two prostate malignant growth cell lines, CWR22Rv1 and TSU-PR1. ABCC10 transcript up-regulated has also been detected in salivary organ adenocarcinoma (Kathawala *et al*., 2014).

In a past report done by Oguri *et al*. (2008), they examined the connection between the cytotoxicity of Paclitaxel and gene expression of ABCC10 using qRT – PCR, demonstrating that the up regulated gene expression levels of ABCC10 and ABCB1 determined in the 17 NSCLC cell line deceased the cytotoxicity of Paclitaxel (Oguri *et al*., 2008).

3.5.3.2. Over Survival analysis *via* PROGGENE on ABCC10

The result obtained from PROGGENE database (Figure 3.4B) shows mRNA gene expression level of ABCC10 gene through examining the over survival for breast cancer patients. From the result obtained, ABCC10 showed a decrease in survival rates for both high and low expression levels of this gene within the 3 years creating a noticeable gap between the two graphs. The plot showed a P value ≤ 0.01 thus the over survival index value is considered to be highly significant and the survival measurement result which makes it a good analysis.

After 5 years, the low expression group of patients (more than 60%) showed over survival of breast cancer. Therefore, high expression of this gene correlates to a potential poor prognostic outcome.

The outcomes from the PROGGENE database is comparable with the outcomes from other studies by Yamada *et al*., (2013) and Estudante *et al*., (2016), which detailed ABCC10 as a hazard factor for breast tumourigenesis, despite the fact that this gene was initially demonstrated to be a hazard factor for advancement of breast malignant growth among Japanese women. The examination uncovered the high expression of ABCC10, which was

increasingly expressed in HER2-enriched, core basal, luminal A BC subtypes ABCC10 expression was also associated with sensitivity and resistance to chemotherapy and poor clinical outcome in breast cancer (Honorat *et al*., 2008).

Another investigation by Honorat *et al.,* (2008) revealed that ABCC10 was up-regulated in ER-α-positive breast tumours, as contrasted with normal breast tissue. Previously, Park *et al.,* (2006) investigated the mRNA levels of ABC transporter genes in breast malignant growth patients who experienced and underwent sequential weekly paclitaxel/(5-FU, epirubicin, and cyclophosphamide) FEC (NAC).

Their examination demonstrated that the efflux of drugs was increased in patients with residual disease as compared with the patients with no pathologic proof of any residual invasive cancer cells in the breast with increased ABCC10 expression (Hlaváč *et al.,* 2013; Elsnerova *et al.,* 2017).

3.5.3.3. Metastasis free survival analysis *via* PROGGENE on ABCC10

PROGGENE database result (Figure 3.4C) shows mRNA expression levels of ABCC10 through examining the metastasis free survival for breast cancer patients. However, the plot from PROGGENE database for ABCC10 shows a decrease in survival rates for both high and low expression levels of this gene within the 3 years. The plot had a P value $\leq$ 0.05 thus making it a good representation of survival measure while being statistically significant.

After 5 years, for the low expression of the gene about 75% of patient' tumours metastasized and reoccurred, therefore high expression of this gene correlates to a poor prognostic outcome. When ABCC10 is expressed at high levels the 5-year non-metastasis free survival rate is about 65%, indicating that after 5 years, only 35% of patients survived and were still part of the study. When ABCC10 is expressed at low levels the 5-year survival rate is over 70%.

ABCB9, ABCB10, and ABCC11 genes, which made up the rest of list 2, produced no results when they were used individually as input in the SurvExpress and PROGGENE databases. The absence of gene expression data may be due to the fact that SurvExpress and PROGGENE databases either have limited information about these genes or may not contain enough experimental information to confirm the expression profiles of these genes. These results also support the idea that these genes could be novel for breast cancer, hence

the databases use the experimental validation as an evidence to implicate the existing genes in different disease processes including cancer.

3.6. Conclusion and summary

The discovery of prognostic and predictive biomarkers in breast cancer research is a first priority and concern. It is thus of importance to validate prognostic or predictive value of candidate genes in breast cancer populations.

In this chapter, the prioritized list of genes was analysed for their survival, over survival, prognostic and predictive metastasis free survival value for breast cancer using several *in silico* tools.

For the prioritized genes (List 2), ABCB9, ABCB10, and ABCC12 displayed no results on the SurvExpress and PROGGENE databases. These results support the idea that these genes maybe novel for their association with breast cancer. Hence, the databases uses the experimental validation as evidence to implicate the existing genes in different disease processes including cancer.

For the survival and prognostic evaluation of the prioritized genes, five out of the 10 genes showed a fair to very good prognostic value for breast cancer based on the expression values of these genes within high- and low-risk groups for breast cancer survival. The results from this chapter prioritize these genes even further for future examination through various other *in silico* tools as well as molecular validation since these genes also showed association with other cancers as well.

In this chapter, some significant genes that could autonomously fill in as possible breast cancer diagnosis, prognosis, or predictive biomarkers were identified by utilizing the PROGENE V2 and SurvExpress biomarker validation tools. However, the difference in important genes distinguished from two distinctive datasets from the SurvExpress biomarker analytical tool, propose that a board of different genes will be progressively suitable as applicant biomarkers for breast cancer than the individual genes. The outcomes affirmed that the vast majority of the distinguished genes are differentially expressed in breast malignant tumour and the expression level of these genes could fill in as prognostic or predictive indicators biomarkers in breast cancer.

CHAPTER FOUR: General conclusions

4.1 Conclusion and future work

Breast cancer remains the most frequent type of cancer amongst women and a global health problem (Ahmedin *et al.*, 2011) Based on the challenges posed by breast cancer, there is an urgent need for biomarkers to improve the diagnosis, prognosis, and management, as well as monitoring patient response to therapeutics. Chemotherapy resistance is a critical issue in the management of breast cancer patients. Breast cancer as a disease represents one of the best models to study the functionality of the ABC drug transporters family. The effluxing of Tamoxifen caused by the ABC transporter genes leads to ineffective therapeutic intervention in breast cancer and result in poor prognosis and increased mortality rates (Heel *et al*., 1978; Ruddy *et al*., 2013).

Ferlay *et al*. (2015), reported that an estimated 1.67 million women had been diagnosed with breast cancer worldwide. According to the NCR 2011, breast cancer 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). The advent of high throughput analysis technologies have delivered a faster means of genomic analysis, which have provided differential expression levels of gene/s in a disease condition and whose protein products have been predicted and published as candidate cancer biomarkers (Chatterjee and Zetter, 2005; Khalid Raza, 2012; Kunin *et al*., 2008).

The *in silico* approach could be a more effective strategy for biomarker discovery and validation, which is considered faster and a more cost effective process compared to the traditional biomarker discovery and validation methods.

The need for new biomarkers to facilitate the processes of diagnosis and prognosis of breast cancer has become an important step in breast cancer management and treatment and the ABC transporter genes has emerged as worthwhile to investigate for this purpose. The ABC transporter genes has shown potential use as diagnostic, prognostic and therapeutic resolutions for different cancers including breast cancer (Sharom, 2008). Therefore, the main aim of this study was to identify, characterize, and prioritize novel ABC transporter genes involved with Tamoxifen MDR in breast cancer patients for the successful treatment and management of breast cancer using several *in silico* methods.

4.1.2. Chapter Two: Categorization, prioritization through gene expression and functional network interaction analysis of the ABC transporter genes implicated in Tamoxifen resistance

The aim of this chapter was to categorize and prioritize ABC transporter genes as prognostic and predictive biomarkers for breast cancer treatment especially for patients who are using Tamoxifen as a chemotherapeutic regiment.

Firstly, the literature was used to identify all of the ABC transporter genes that are associated with breast cancer. This analysis resulted in 48 genes being identified, which was shortlisted and prioritized to 10 genes using the following inclusive criteria: (i) expression in breast tissue, (ii) expression in breast cancer tissue, (iii) involvement in drug transportation and (iv) involvement in multi-drug resistance to Tamoxifen. The HPA and TiGER databases were employed to generate the shortlisted number of genes. The analysis using HPA and TiGER showed that ABCC10 are overexpressed in breast cancer tissue. This observation was in line with the study of Park *et al*. (2006), who showed the up-regulation of this gene in patients with residual breast cancer. The study by Zhang *et al*., 2016, showed that several compounds interacted with ABCC10 and that this gene can confer resistance to these compounds by decreasing their bioavailability, including Tamoxifen (Balaji *et al*., 2016a). ABCC11 was also showed to be overexpressed in breast cancer using HPA in this study with this observation corroborated by several other studies that showed this gene's involvement in confering resistance to several chemical compounds in breast cancer tissue (Toyoda and Ishikawa, 2011).

An interaction network was drawn using STRING with the 10 shortlisted genes as input. STRING analysis showed all of the shortlisted genes interacting with each other through experimental evidence, co-expression and evidence from the literature.

The 10 genes were further analysed using intersectional Venn diagrams to examine the toxicity relation between the identified genes and Tamoxifen as the desired chemotherapy of interaction. Based on this analysis, five genes showed intersectional interaction *via* the CTD, using "MyGeneVenn" showing a relationship between the genes and Tamoxifen based on toxicity. The prioritized genes list 2 were further analysed and validated for their progression and survival in breast cancer patients using different bioinformatics tools in the subsequent chapter.

4.1.3. Chapter Three: Survival, Prognostic, and Predictive analysis of the prioritized ABC transporter genes for breast cancer management using *in silico* approaches

The aim in this chapter was to evaluate the survival, prognostic and predictive value for the prioritized list of ABC transporter genes using several prognostics software and databases.

The SurvExpress and PROGGENE were used to analyse the prognostic ability, over survival and metastasis free survival of the individual candidate biomarkers based on differentiating between the low and the high risk patient's groups based on their expression levels. From the survival curves, ABCC4, ABCC5, and ABCC10 showed significant statistical values in predicting the prognostic outcome of breast cancer based on the differential expression value of the biomarkers.

Interestingly, ABCB9, ABCB11 and ABCC11 genes obtain no result from the SurvExpress and PROGGENE databases on their prognostic value to implicate these genes in breast cancer. However, this results support the notion that these genes could be novel for breast cancer since the databases used are based on experimental evidence to implicate existing genes in different disease processes, including cancer.

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

Results obtained from the PROGGENE survival and predictive analysis showed good prognostic values for the genes ABCB2, ABCC4, ABCC5, ABCC10 and ABCC12 with their significance measured by the probability value (*Pv*) (0.053, 0.001118, 0.01286, 0.00604, 0.00157 respectively).

The variance in the set of genes being significant for use as prognostic markers, suggest that this set of genes will be more effective when combined than used as individual biomarkers for breast cancer management. The set of genes showed positive prognostic

and predictive values using the two different datasets obtained from the SurvExpress database.

The difference in result outcomes from the different bioinformatics tools used can be a consequence of different cancer types and techniques used in analyzing the differential gene expression of database entries. The ability of these bioinformatics tools to confirm the same genes as having value as potential biomarkers, is an indication that these tools are effective for *in silico* analysis of prospective biomarkers.

4.2. Future Work

Future work would include examination of the expression profiles of the ABC prioritized genes in a larger panel of cells and tissues. It would also be interesting to analyse the expression level of these up-regulated protein, using quantitative immunofluorescence analysis or Enzyme-Linked Immunosorbent Assay.

Also, as these candidate proteins are annotated to be cell surface proteins and are expected to be shed into bodily fluids, the presence of these proteins need to be validated in various bodily fluids and tissue samples of breast cancer patients.

Furthermore, to investigate the 3'- 5' untranslated region by analyzing the promoter content and identify any mutation in the promoter region of these prioritized genes and their relation in breast cancer chemotherapy resistance using Cape Analysis of Gene Expression (CAGE). CAGE is a gene expression technique which has been used in molecular biology laboratory to produce a snapshot of the 3'- 5′ Untranslated region end of the messenger RNA population in a biological sample. This will be followed by Single nucleotide polymorphisms profiling in populations of the most suitable candidate biomarkers to map its frequency in chemotherapy resistance in South African breast cancer cases. Lastly, reverse engineer an analogue of the most suitable candidate biomarker that can preferentially bind these efflux pumps instead of Tamoxifen, which will insure that the chemotherapeutic agent remains at very high concentration within breast cancer cells as to increase its efficacy.

References

ACS, 2019. American Cancer Society (ACS).

Aguirre-Gamboa, R., Gomez-Rueda, H., Martínez-Ledesma, E., Martínez-Torteya, A., Chacolla-Huaringa, R., Rodriguez-Barrientos, A., Tamez-Peña, J.G., Treviño, V., 2013. SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. PLoS One 8, 1–9. https://doi.org/10.1371/journal.pone.0074250.

Ahmedin J, Bray F, Center M, Ferlay J, Ward E, F.D., 2011. Global cancer statistics. CA Cancer J Clin. 2011;, in: Ca Cancer J Clin. pp. 69–90. https://doi.org/10.3322/caac.20107.Available.

Al-Mansouri, L.J., Alokail, M.S., 2006. Molecular basis of breast cancer. Saudi Med. J. 27, 9–16. https://doi.org/10.1007/978-3-642-18736-0.

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schnitt, S., Sellers, W.R., Polyak, K., 2004. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 6, 17–32. https://doi.org/10.1016/j.ccr.2004.06.010.

Amatori, S., Persico, G., Fanelli, M., 2017. Real-time quantitative PCR array to study drug-induced changes of gene expression in tumor cell lines. J. Cancer Metastasis Treat. 3, 90. https://doi.org/10.20517/2394-4722.2017.22.

Arps, D.P., Healy, P., Zhao, L., Kleer, C.G., 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 Res. Treat. 138, 719–726. https://doi.org/10.1007/s10549-013-2493-2.

Baak, J.P.A., Janssen, E.A.M., Soreide, K., Heikkilæ, R., 2005. Genomics and proteomics - The way forward. Ann. Oncol. 16, 30–44. https://doi.org/10.1093/annonc/mdi728.

Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R., Palacios, J., Rakha, E.A., Richardson, A.L., Schmitt, F.C., Tan, P.H., Tse, G.M., Weigelt, B., Ellis, I.O., Reis-Filho, J.S., 2011. Basal-like and

triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. Mod. Pathol. 24, 157–167. https://doi.org/10.1038/modpathol.2010.200.

Bagnoli, M., Beretta, G.L., Gatti, L., Pilotti, S., Alberti, P., Tarantino, E., Barbareschi, M., Canevari, S., Mezzanzanica, D., Perego, P., 2013. Clinicopathological impact of ABCC1/MRP1 and ABCC4/MRP4 in epithelial ovarian carcinoma. Biomed Res. Int. 2013. https://doi.org/10.1155/2013/143202.

Balaji, S.A., Udupa, N., Chamallamudi, M.R., Gupta, V., Rangarajan, A., 2016a. Role of the drug transporter ABCC3 in breast cancer chemoresistance. PLoS One. https://doi.org/10.1371/journal.pone.0155013.

Balaji, S.A., Udupa, N., Chamallamudi, M.R., Gupta, V., Rangarajan, A., 2016b. Role of the drug transporter ABCC3 in breast cancer chemoresistance. PLoS One 11, 1–22. https://doi.org/10.1371/journal.pone.0155013.

Barabási, A.L., Oltvai, Z.N., 2004. Network biology: Understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113. https://doi.org/10.1038/nrg1272.

Baxevanis, A.D., Francis Ouellette, B.F., 2004. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins., 3rd ed, Annals of Biomedical Engineering. https://doi.org/10.1007/s10439-006-9105-9.

Begicevic, R.R., Falasca, M., 2017. ABC transporters in cancer stem cells: Beyond chemoresistance. Int. J. Mol. Sci. 18. https://doi.org/10.3390/ijms18112362.

Benjeddou, M., 2010. Solute carrier transporters: Pharmacogenomics research opportunities in Africa. African J. Biotechnol. 9, 9191–9195. https://doi.org/10.5897/AJB2010.000-3343.

Bera, T.K., Iavarone, C., Kumar, V., Lee, S., Lee, B., Pastan, I., 2002. MRP9, an unusual truncated member of the ABC transporter superfamily, is highly expressed in breast cancer. Proc. Natl. Acad. Sci. U. S. A. 99, 6997–7002. https://doi.org/10.1073/pnas.102187299.

Bera, T.K., Lee, S., Salvatore, G., Lee, B., Pastan, I., 2001. MRP8, A New Member of ABC Transporter Superfamily, Identified by EST Database Mining and Gene Prediction Program, Is Highly Expressed in Breast Cancer. Mol. Med. 7, 409–416.

https://doi.org/10.1007/bf03401856.

Bertoncini, S., Blanco-Rojo, R., Baeza, C., Arroyo-Pardo, E., Vaquero, M.P., López-Parra, A.M., 2011. A novel snapshot assay to detect genetic mutations related to iron metabolism. Genet. Test. Mol. Biomarkers 15, 173–179. https://doi.org/10.1089/gtmb.2010.0140.

Bhatt, A.N., Mathur, R., Farooque, A., Verma, A., Dwarakanath, B.S., 2010. Cancer biomarkers - Current perspectives. Indian J. Med. Res. 132, 129–149.

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., Tavaré, S., Caldas, C., Miska, E.A., 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol. 8. https://doi.org/10.1186/gb-2007-8-10-r214.

Borel, F., Han, R., Visser, A., Petry, H., van Deventer, S.J.H., Jansen, P.L.M., Konstantinova, P., 2012. Adenosine triphosphate-binding cassette transporter genes up-regulation in untreated hepatocellular carcinoma is mediated by cellular microRNAs. Hepatology 55, 821–832. https://doi.org/10.1002/hep.24682.

Bravaccini, S., Granato, A.M., Medri, L., Foca, F., Falcini, F., Zoli, W., Ricci, M., Lanzanova, G., Masalu, N., Serra, L., Buggi, F., Folli, S., Silvestrini, R., Amadori, D., 2013. Biofunctional characteristics of in situ and invasive breast carcinoma. Cell. Oncol. 36, 303–310. https://doi.org/10.1007/s13402-013-0135-7.

Brockmöller, J., Tzvetkov, M. V., 2008. Pharmacogenetics: Data, concepts and tools to improve drug discovery and drug treatment. Eur. J. Clin. Pharmacol. https://doi.org/10.1007/s00228-007-0424.

Butti, R., Gunasekaran, V.P., Kumar, T.V.S., Banerjee, P., Kundu, G.C., 2019. Breast cancer stem cells: Biology and therapeutic implications. Int. J. Biochem. Cell Biol. 107, 38–52. https://doi.org/10.1016/j.biocel.2018.12.001.

Callaghan, R., Higgins, C.F., 1995. Interaction of tamoxifen with the multidrug resistance P-glycoprotein. Br. J. Cancer 71, 294–299. https://doi.org/10.1038/bjc.1995.59.

Carlomagno, N., Incollingo, P., Tammaro, V., Peluso, G., Rupealta, N., Chiacchio, G., Sandoval Sotelo, M.L., Minieri, G., Pisani, A., Riccio, E., Sabbatini, M., Bracale, U.M.,

Calogero, A., Dodaro, C.A., Santangelo, M., 2017. Diagnostic, Predictive, Prognostic, and Therapeutic Molecular Biomarkers in Third Millennium: A Breakthrough in Gastric Cancer. Biomed Res. Int. 2017. https://doi.org/10.1155/2017/7869802.

Chang, M., 2012. Tamoxifen resistance in breast cancer. Biomol. Ther. 20, 256–267. https://doi.org/10.4062/biomolther.2012.20.3.256.

Chatterjee, S.K., Zetter, B.R., 2005. Cancer biomarkers: Knowing the present and predicting the future. Futur. Oncol. 1, 37–50. https://doi.org/10.1517/14796694.1.1.37.

Chen, Z., Shi, T., Zhang, L., Zhu, P., Deng, M., Huang, C., Hu, T., Jiang, L., Li, J., 2016. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade. Cancer Lett. 370, 153–164. https://doi.org/10.1016/j.canlet.2015.10.010.

Choudhuri, S., 2014. The Beginning of Bioinformatics. Bioinforma. Beginners 73–76. https://doi.org/10.1016/b978-0-12-410471-6.00004-9.

Choudhuri, S., Klaassen, C., 2006. Structure, function, expression, genomic organization, and single nucleotide polymorphisms of human ABCB1 (MDR1), ABCC (MRP), and ABCG2 (BCRP) efflux transporters. Int. J. Toxicol. 25, 231–259. https://doi.org/10.1080/10915810600746023.

Clemons, M., Danson, S., Howell, A., 2002. Tamoxifen ('Nolvadex'): A review. Cancer Treat. Rev. 28, 165–180. https://doi.org/10.1016/S0305-7372(02)00036.

Cocca, M., Bedognetti, D., Bianca, M., Gasparini, P., Girotto, G., 2016. Pharmacogenetics driving personalized medicine: Analysis of genetic polymorphisms related to breast cancer medications in Italian isolated populations. J. Transl. Med. 14, 1–12. https://doi.org/10.1186/s12967-016-0778.

Colomer, R., Aranda-López, I., Albanell, J., García-Caballero, T., Ciruelos, E., López-García, M., Cortés, J., Rojo, F., Martín, M., Palacios-Calvo, J., 2018. Biomarkers in breast cancer: A consensus statement by the Spanish Society of Medical Oncology and the Spanish Society of Pathology. Clin. Transl. Oncol. 20, 815–826. https://doi.org/10.1007/s12094-017-1800-5.

Cook, A.C., Tuck, A.B., McCarthy, S., Turner, J.G., Irby, R.B., Bloom, G.C., Yeatman, T.J.,

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. Mol. Carcinog. 43, 225–236. https://doi.org/10.1002/mc.20105.

Craig Venter, J., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Yuan Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M.H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Lai Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Ni Tint, N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J.,

Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Deslattes Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science (80-.). 291, 1304–1351. https://doi.org/10.1126/science.1058040.

Criscitiello, C., Fumagalli, D., Saini, K.S., Loi, S., 2011. Tamoxifen in early-stage estrogen receptorpositive breast cancer: Overview of clinical use and molecular biomarkers for patient selection. Onco. Targets. Ther. 4, 1–11. https://doi.org/10.2147/OTT.S10155.

Crowgey, E.L., Matlock, A., Venkatraman, V., Fert-, J., Eyk, J.E. Van, 2018. Data to Knowledge Pipelines. https://doi.org/10.1007/978-1-4939-6783-4.

Daugelaite, J., O' Driscoll, A., Sleator, R.D., 2013. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. ISRN Biomath. 2013, 1–14. https://doi.org/10.1155/2013/615630.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., McMorran, R., Wiegers, J., Wiegers, T.C., Mattingly, C.J., 2019. The Comparative Toxicogenomics Database: Update 2019. Nucleic Acids Res. 47, D948–D954. https://doi.org/10.1093/nar/gky868.

Debouck, C., 2009. Integrating genomics across drug discovery and development. Toxicol. Lett. 186, 9–12. https://doi.org/10.1016/j.toxlet.2008.09.011.

DeVita, V.T., Chu, E., 2008. A history of cancer chemotherapy. Cancer Res. 68, 8643–8653. https://doi.org/10.1158/0008-5472.CAN-07-6611.

Domenichini, A., Adamska, A., Falasca, M., 2019. ABC transporters as cancer drivers: Potential functions in cancer development. Biochim. Biophys. Acta - Gen. Subj. 1863, 52–60. https://doi.org/10.1016/j.bbagen.2018.09.019.

Donepudi, M.S., Kondapalli, K., Amos, S.J., Venkanteshan, P., 2014. Breast cancer statistics and markers. J. Cancer Res. Ther. 10, 506–511. https://doi.org/10.4103/0973-

1482.137927.

Dowsett, M., Houghton, J., Iden, C., Salter, J., Farndon, J., A'Hern, R., Sainsbury, R., Baum, M., 2006. Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status. Ann. Oncol. 17, 818–826. https://doi.org/10.1093/annonc/mdl016.

Doyle, L.A., Ross, D.D., 2003. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). Oncogene 22, 7340–7358. https://doi.org/10.1038/sj.onc.1206938.

Duffy, M.J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., Cardoso, F., 2017. Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). Eur. J. Cancer 75, 284–298. https://doi.org/10.1016/j.ejca.2017.01.017.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.L., Bogdanova, N., Schürmann, P., Dörk, T., Tollenaar, R.A.E.M., Jacobi, C.E., Devilee, P., Klijn, J.G.M., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W.R., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., Van Den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.D., Spurdle, A.B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V.M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., Ponder, B.A.J., Luccarini, C., Conroy, D., Shah, M., Munday, H., Jordan, C., Perkins, B., West, J., Redman, K., Driver, K., Aghmesheh, M., Amor, D., Andrews, L., Antill, Y., Armes, J., Armitage, S., Arnold, L., Balleine, R., Begley, G., Beilby, J., Bennett, I., Bennett, B., Berry, G., Blackburn, A., Brennan, M., Brown, M., Buckley, M.,

Burke, J., Butow, P., Byron, K., Callen, D., Campbell, I., Clarke, C., Colley, A., Cotton, D., Cui, J., Culling, B., Cummings, M., Dawson, S.J., Dixon, J., Dobrovic, A., Dudding, T., Edkins, T., Eisenbruch, M., Farshid, G., Fawcett, S., Field, M., Firgaira, F., Fleming, J., Forbes, J., Friedlander, M., Gaff, C., Gardner, M., Gattas, M., George, P., Gill, G., Gill, G., Goldblatt, J., Greening, S., Haan, E., Haan, E., Harris, M., Hart, S., Hayward, N., Hopper, J., Humphrey, E., Jenkins, M., Kefford, R., Kefford, R., Kirk, J., Kollias, J., Kovalenko, S., Lakhani, S., Leary, J., Lim, J., Lindeman, G., Lipton, L., Lobb, L., Maclurcan, M., Marsh, D., Marsh, D., McKay, M., Anne McLachlan, S., Milne, R., Mitchell, G., Mitchell, G., Newman, B., O'Loughlin, I., Osborne, R., Peters, L., Price, M., Price, M., Reeve, J., Reeve, T., Richards, R., Rinehart, G., Robinson, B., Rudzki, B., Salisbury, E., Saunders, C., Saunders, C., Scott, E., Scott, E., Seshadri, R., Shelling, A., Shelling, A., Suthers, G., Suthers, G., Taylor, D., Tennant, C., Townshend, S., Townshend, S., Tyler, J., Venter, D., Venter, D., Visvader, J., Walpole, I., Ward, R., Warner, B., Warner, B., Warren, G., Watson, E., Williams, R., Wilson, J., Winship, I., Young, M.A., Bowtell, D., DeFazio, A., DeFazio, A., Gertig, D., Webb, P., 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087–1093. https://doi.org/10.1038/nature05887.

Eckford, P.D.W., Sharom, F.J., 2009. ABC efflux pump-based resistance to chemotherapy drugs. Chem. Rev. 109, 2989–3011. https://doi.org/10.1021/cr9000226.

El-Awady, R., Saleh, E., Hashim, A., Soliman, N., Dallah, A., Elrasheed, A., Elakraa, G., 2017. The role of eukaryotic and prokaryotic ABC transporter family in failure of chemotherapy. Front. Pharmacol. 7, 1–15. https://doi.org/10.3389/fphar.2016.00535.

Elsnerova, K., Bartakova, A., Tihlarik, J., Bouda, J., Rob, L., Skapa, P., Hruda, M., Gut, I., Mohelnikova-Duchonova, B., Soucek, P., Vaclavikova, R., 2017. Gene expression profiling reveals novel candidate markers of ovarian carcinoma intraperitoneal metastasis. J. Cancer 8, 3598–3606. https://doi.org/10.7150/jca.20766.

Estudante, M., Soveral, G., Morais, J.G., Benet, L.Z., 2016. Insights into solute carriers: Physiological functions and implications in disease and pharmacokinetics. Medchemcomm 7, 1462–1478. https://doi.org/10.1039/c6md00188b.

Fanale, D., Amodeo, V., Corsini, L.R., Rizzo, S., Bazan, V., Russo, A., 2012. Breast cancer genome-wide association studies: There is strength in numbers. Oncogene 31, 2121–2128.

https://doi.org/10.1038/onc.2011.408.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide : Sources , methods and major patterns in GLOBOCAN 2012 386. https://doi.org/10.1002/ijc.29210.

Fletcher, J.I., Haber, M., Henderson, M.J., Norris, M.D., 2010. ABC transporters in cancer: More than just drug efflux pumps. Nat. Rev. Cancer 10, 147–156. https://doi.org/10.1038/nrc2789.

Fletcher, J.I., Williams, R.T., Henderson, M.J., Norris, M.D., Haber, M., 2016. ABC transporters as mediators of drug resistance and contributors to cancer cell biology. Drug Resist. Updat. 26, 1–9. https://doi.org/10.1016/j.drup.2016.03.001.

Fraguas-Sánchez, A.I., Martín-Sabroso, C., Fernández-Carballido, A., Torres-Suárez, A.I., 2019. Current status of nanomedicine in the chemotherapy of breast cancer. Cancer Chemother. Pharmacol. 84, 689–706. https://doi.org/10.1007/s00280-019-03910.

França, L.K.L., Bitencourt, A.G.V., Paiva, H.L.S., Silva, C.B., Pereira, N.P., Paludo, J., Graziano, L., Guatelli, C.S., de Souza, J.A., Marques, E.F., 2017. Papel da ressonância magnética no planejamento terapêutico das pacientes com câncer de mama: Comparação com exames convencionais. Radiol. Bras. 50, 76–81. https://doi.org/10.1590/0100-3984.2015.0124.

Frances, A., 2013. The past, present and future of psychiatric diagnosis. World Psychiatry 12, 111–112. https://doi.org/10.1002/wps.20027.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., Jensen, L.J., 2013. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, 808–815. https://doi.org/10.1093/nar/gks1094.

Gaudet, M.M., Gapstur, S.M., Sun, J., Ryan Diver, W., Hannan, L.M., Thun, M.J., 2013. Active smoking and breast cancer risk: Original cohort data and meta-analysis. J. Natl. Cancer Inst. 105, 515–525. https://doi.org/10.1093/jnci/djt023.

Ginsburg, G.S., Willard, H.F., 2009. Genomic and personalized medicine: foundations and applications. Transl. Res. 154, 277–287. https://doi.org/10.1016/j.trsl.2009.09.005.

Goswami, C.P., Nakshatri, H., 2013. PROGgene: Gene expression based survival analysis web application for multiple cancers. J. Clin. Bioinforma. 3, 1–9. https://doi.org/10.1186/2043-9113-3-22.

Grumolato, L., Aaronson, S.A., 2014. Oncogenes and Signal Transduction. Mol. Basis Cancer Fourth Ed. 19–34.e3. https://doi.org/10.1016/B978-1-4557-4066-6.00002.

Gutschner, T., Diederichs, S., 2012. The hallmarks of cancer: A long non-coding RNA point of view. RNA Biol. 9, 703–709. https://doi.org/10.4161/rna.20481.

Györffy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., 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 Res. Treat. 123, 725–731. https://doi.org/10.1007/s10549-009-0674.

Hahn, W.C., Weinberg, R.A., 2014. Cancer: A Genetic Disorder. Mol. Basis Cancer Fourth Ed. 3–18.e1. https://doi.org/10.1016/B978-1-4557-4066-6.00001.

Hamed, A.R., Abdel-Azim, N.S., Shams, K.A., Hammouda, F.M., 2019. Targeting multidrug resistance in cancer by natural chemosensitizers. Bull. Natl. Res. Cent. 43. https://doi.org/10.1186/s42269-019-0043.

Hamilton, a, Piccart, M., 2000. Review the treatment of breast cancer : A review of the literature on HER-2. Ann. Oncol. 11, 647–663.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. Cell 144, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

Hanahan, D., Weinberg, R.A., 2000. The Hallmarks of cancer. Cell Press 100, 57–70. https://doi.org/10.1007/BF03091804.

Hassiotou, F., Geddes, D., 2013. Anatomy of the human mammary gland: Current status of knowledge. Clin. Anat. 26, 29–48. https://doi.org/10.1002/ca.22165.

Hazlehurst, L., Hacker, M., 2009. Drug Resistance, 1st ed, Pharmacology. Elsevier Inc. https://doi.org/10.1016/B978-0-12-369521-5.00015.

Heel, R.C., Brogden, R.N., Speight, T.M., Avery, G.S., 1978. Tamoxifen: A Review of its Pharmacological Properties and Therapeutic Use in the Treatment of Breast Cancer.

Drugs 16, 1–24. https://doi.org/10.2165/00003495-197816010-00001.

Hlaváč, V., Brynychová, V., Václavíková, R., Ehrlichová, M., Vrána, D., Pecha, V., Koževnikovová, R., Trnková, M., Gatěk, J., Kopperová, D., Gut, I., Souček, P., 2013. The expression profile of ATP-binding cassette transporter genes in breast carcinoma. Pharmacogenomics 14, 515–529. https://doi.org/10.2217/pgs.13.26.

Hlavata, I., Mohelnikova-Duchonova, B., Vaclavikova, R., Liska, V., Pitule, P., Novak, P., Bruha, J., Vycital, O., Holubec, L., Treska, V., Vodicka, P., Soucek, P., 2012. The role of ABC transporters in progression and clinical outcome of colorectal cancer. Mutagenesis 27, 187–196. https://doi.org/10.1093/mutage/ger075.

Honorat, M., Mesnier, A., Vendrell, J., Guitton, J., Bieche, I., Lidereau, R., Kruh, G.D., Dumontet, C., Cohen, P., Payen, L., 2008. ABCC11 expression is regulated by estrogen in MCF7 cells, correlated with estrogen receptor α expression in postmenopausal breast tumors and overexpressed in tamoxifen-resistant breast cancer cells. Endocr. Relat. Cancer 15, 125–138. https://doi.org/10.1677/ERC-07-0189.

Hopper-Borge, E., Chen, Z.S., Shchaveleva, I., Belinsky, M.G., Kruh, G.D., 2004. Analysis of the drug resistance profile of multidrug resistance protein 7 (ABCC10): Resistance to docetaxel. Cancer Res. 64, 4927–4930. https://doi.org/10.1158/0008-5472.CAN-03-3111.

Hulka, B.S., Moorman, P.G., 2001. Breast cancer: Hormones and other risk factors. Maturitas 38, 103–113. https://doi.org/10.1016/S0378-5122(00)00196.

Iorio, M. V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Ménard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M., Croce, C.M., 2005. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 65, 7065–7070. https://doi.org/10.1158/0008-5472.CAN-05-1783.

Ishikawa, T., Toyoda, Y., Yoshiura, K.I., Niikawa, N., 2013. Pharmacogenetics of human ABC transporter ABCC11: New insights into apocrine gland growth and metabolite secretion. Front. Genet. 3, 1–13. https://doi.org/10.3389/fgene.2012.00306.

Jemal, A., Bray, F., Forman, D., O'Brien, M., Ferlay, J., Center, M., Parkin, D.M., 2012. Cancer burden in Africa and opportunities for prevention. Cancer 118, 4372–4384.

https://doi.org/10.1002/cncr.27410.

Jones, P.M., George, A.M., 2004. The ABC transporter structure and mechanism: Perspectives on recent research. Cell. Mol. Life Sci. 61, 682–699. https://doi.org/10.1007/s00018-003-3336.

Jurca, G., Addam, O., Aksac, A., Gao, S., Özyer, T., Demetrick, D., Alhajj, R., 2016. Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends. BMC Res. Notes 9, 1–35. https://doi.org/10.1186/s13104-016-2023.

Karakach, T.K., Flight, R.M., Douglas, S.E., Wentzell, P.D., 2010. An introduction to DNA microarrays for gene expression analysis. Chemom. Intell. Lab. Syst. 104, 28–52. https://doi.org/10.1016/j.chemolab.2010.04.003.

Karvar, S., 2014. The role of ABC transporters in anticancer drug transport. Turkish J. Biol. 38, 800–805. https://doi.org/10.3906/biy-1407.

Kathawala, R.J., Gupta, P., Ashby, C.R., Chen, Z.S., 2015. The modulation of ABC transporter-mediated multidrug resistance in cancer: A review of the past decade. Drug Resist. Updat. 18, 1–17. https://doi.org/10.1016/j.drup.2014.11.002.

Kathawala, R.J., Wang, Y.J., Ashby, C.R., Chen, Z.S., 2014. Recent advances regarding the role of ABC subfamily C member 10 (ABCC10) in the efflux of antitumor drugs. Chin. J. Cancer 33, 223–230. https://doi.org/10.5732/cjc.013.10122.

Kawaguchi, K., Morita, M., 2016. ABC Transporter Subfamily D: Distinct Differences in Behavior between ABCD1-3 and ABCD4 in Subcellular Localization, Function, and Human Disease. Biomed Res. Int. 2016. https://doi.org/10.1155/2016/6786245.

Khalid Raza, 2012. Application of data mining in bioinformatics. indian J. Comput. Sci. Eng. 1, 114–118. https://doi.org/10.4028/www.scientific.net/amm.0.1776.

Khalilpour, A., Kilic, T., Khalilpour, S., Álvarez, M.M., Yazdi, I.K., 2017. Proteomic-based biomarker discovery for development of next generation diagnostics. Appl. Microbiol. Biotechnol. 101, 475–491. https://doi.org/10.1007/s00253-016-8029.

Koepsell, H., Lips, K., Volk, C., 2007. Polyspecific organic cation transporters: Structure, function, physiological roles, and biopharmaceutical implications. Pharm. Res. 24, 1227–1251. https://doi.org/10.1007/s11095-007-9254.

Kolesnikova, M., Sen'Kova, A., Tairova, S., Ovchinnikov, V., Pospelova, T., Zenkova, M., 2019. Clinical and prognostic significance of cell sensitivity to chemotherapy detected in vitro on treatment response and survival of leukemia patients. J. Pers. Med. 9. https://doi.org/10.3390/jpm9020024.

Kontos, C.K., Adamopoulos, P.G., Scorilas, A., 2015. Prognostic and predictive biomarkers in prostate cancer. Expert Rev. Mol. Diagn. 15, 1567–1576. https://doi.org/10.1586/14737159.2015.1110022.

Kotsopoulos, J., Huzarski, T., Gronwald, J., Singer, C.F., Moller, P., Lynch, H.T., Armel, S., Karlan, B., Foulkes, W.D., Neuhausen, S.L., Senter, L., Tung, N., Weitzel, J.N., Eisen, A., Metcalfe, K., Eng, C., Pal, T., Evans, G., Sun, P., Lubinski, J., Narod, S.A., 2017. Bilateral oophorectomy and breast cancer risk in BRCA1 and BRCA2 mutation carriers. J. Natl. Cancer Inst. 109, 1–7. https://doi.org/10.1093/jnci/djw177.

Kruh, G.D., Guo, Y., Hopper-Borge, E., Belinsky, M.G., Chen, Z.S., 2007. ABCC10, ABCC11, and ABCC12. Pflugers Arch. Eur. J. Physiol. 453, 675–684. https://doi.org/10.1007/s00424-006-0114-1.

Kulasingam, V., Diamandis, E.P., 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat. Clin. Pract. Oncol. 5, 588–599. https://doi.org/10.1038/ncponc1187.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., Hugenholtz, P., 2008. A Bioinformatician's Guide to Metagenomics. Microbiol. Mol. Biol. Rev. https://doi.org/10.1128/mmbr.00009-08.

Kyle Strimbu, Tavel, J.A., M.D., 2010. What are Biomarkers? PMC November, 463–466. https://doi.org/10.1097/COH.0b013e32833ed177.

Lal, S., Sutiman, N., Ooi, L.L., Wong, Z.W., Wong, N.S., Ang, P.C.S., Chowbay, B., 2017. Pharmacogenetics of ABCB5, ABCC5 and RLIP76 and doxorubicin pharmacokinetics in Asian breast cancer patients. Pharmacogenomics J. 17, 337–343. https://doi.org/10.1038/tpj.2016.17.

Leabman, M.K., Huang, C.C., DeYoung, J., Carlson, E.J., Taylor, T.R., De la Cruz, M., Johns, S.J., Stryke, D., Kawamoto, M., Urban, T.J., Kroetz, D.L., Ferrin, T.E., Clark, A.G., Risch, N., Herskowitz, I., Giacomini, K.M., Altman, R.B., Benet, L.Z., Brett, C.M.,

Burchard, E.G., Castro, R., Edwards, R.H., Gitschier, J., Glatt, C.E., Ho, C., Klein, T., Lin, E.T., Reus, V.I., Sadee, W., Salazar, M., Schaefer, C., Sheiner, L.B., Tran, C., Vulpe, C., Wright, E.M., 2003. Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. Proc. Natl. Acad. Sci. U. S. A. 100, 5896–5901. https://doi.org/10.1073/pnas.0730857100.

Leonard, G.D., 2003. The Role of ABC Transporters in Clinical Practice. Oncologist 8, 411–424. https://doi.org/10.1634/theoncologist.8-5-411.

Levine, A.J., Hu, W., Feng, Z., n.d. Ch. 03 Tumor Suppressor Genes.

Li, W., Zhang, H., Assaraf, Y.G., Zhao, K., Xu, X., Xie, J., Yang, D.H., Chen, Z.S., 2016. Overcoming ABC transporter-mediated multidrug resistance: Molecular mechanisms and novel therapeutic drug strategies. Drug Resist. Updat. 27, 14–29. https://doi.org/10.1016/j.drup.2016.05.001.

Liang, P., 2003. Bioinformatics for Cancer Research. Bioinformatics.

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., Ooi, E.J.J., Nam, R.K., Allen, P.D., Steele, G.S., Wassmann, K., Richie, J.P., Liew, C.C., 2012. Blood-Based Biomarkers of Aggressive Prostate Cancer. PLoS One 7, 1–7. https://doi.org/10.1371/journal.pone.0045802.

Litviakov, N. V., Cherdyntseva, N. V., Tsyganov, M.M., Slonimskaya, Y.M., Ibragimova, M.K., Kazantseva, P. V., Kzhyshkowska, J., Choinzonov, E.L., 2016. Deletions of multidrug resistance gene loci in breast cancer leads to the down-regulation of its expression and predict tumor response to neoadjuvant chemotherapy. Oncotarget 7, 7829–7841. https://doi.org/10.18632/oncotarget.6953.

Liu, R., Guo, C.X., Zhou, H.H., 2015. Network-based approach to identify prognostic biomarkers for estrogen receptor-positive breast cancer treatment with tamoxifen. Cancer Biol. Ther. 16, 317–324. https://doi.org/10.1080/15384047.2014.1002360.

Liu, X., Yu, X., Zack, D.J., Zhu, H., Qian, J., 2008. TiGER: A database for tissue-specific gene expression and regulation. BMC Bioinformatics 9, 1–7. https://doi.org/10.1186/1471-2105-9-271.

Lorincz, A.T., Brentnall, A.R., Vasiljevič, N., Scibior-Bentkowska, D., Castanon, A., Fiander,

A., Powell, N., Tristram, A., Cuzick, J., Sasieni, P., 2013. HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. Int. J. Cancer 133, 637–644. https://doi.org/10.1002/ijc.28050.

Lovitt, C.J., Shelper, T.B., Avery, V.M., 2018. Doxorubicin resistance in breast cancer cells is mediated by extracellular matrix proteins. BMC Cancer 18, 1–11. https://doi.org/10.1186/s12885-017-3953-6.

Lowry, K.P., Lee, J.M., Kong, C.Y., McMahon, P.M., Gilmore, M.E., Cott Chubiz, J.E., Pisano, E.D., Gatsonis, C., Ryan, P.D., Ozanne, E.M., Gazelle, G.S., 2012. Annual screening strategies in BRCA1 and BRCA2 gene mutation carriers: A comparative effectiveness analysis. Cancer 118, 2021–2030. https://doi.org/10.1002/cncr.26424.

Ludwig, J.A., Weinstein, J.N., 2005. Biomarkers in cancer staging, prognosis and treatment selection. Nat. Rev. Cancer 5, 845–856. https://doi.org/10.1038/nrc1739.

Maliepaard, M., Scheffer, G.L., Faneyte, I.F., Van Gastelen, M.A., Pijnenborg, A.C.L.M., Schinkel, A.H., Van de Vijver, M.J., Scheper, R.J., Schellens, J.H.M., 2001. Subcellular localization and distribution of the Breast Resistance Protein Transporter in normal human tissues. Cancer Res. 61, 3458–3464.

Mao, Q., Unadkat, J.D., 2005. Role of the breast cancer resistance protein (ABCG2) in drug transport. AAPS J. 7, E118–E133. https://doi.org/10.1208/aapsj070112.

Massarweh, S., Osborne, C.K., Creighton, C.J., Qin, L., Tsimelzon, A., Huang, S., Weiss, H., Rimawi, M., Schiff, R., 2008. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. Cancer Res. 68, 826–833. https://doi.org/10.1158/0008-5472.CAN-07-2707.

Mathers, C.D., Bonita, R., 2009. Current global health status, in: Global Public Health: A New Era. new york, p. 299. https://doi.org/10.1093/acprof:oso/9780199236626.003.0002.

Mattingly, C.J., Colby, G.T., Forrest, J.N., Boyer, J.L., 2003. The Comparative Toxicogenomics Database (CTD). Environ. Health Perspect. 111, 793. https://doi.org/10.1289/ehp.6028.

Mills, D., Gordon, E.J., Casano, A., Lahti, S.M., Nguyen, T., Preston, A., Tondre, J., Wu, K.,

Yanase, T., Chan, H., Chia, D., Esfandiari, M., Himmel, T., Love, S.M., 2011. The physiology of the normal human breast: an exploratory study. J. Physiol. Biochem. 67, 621–627. https://doi.org/10.1007/s13105-011-0109.

Mirtavoos-Mahyari, H., Khosravi, A., 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, 26–34.

Montanari, F., Ecker, G.F., 2015. Prediction of drug-ABC-transporter interaction - Recent advances and future challenges. Adv. Drug Deliv. Rev. 86, 17–26. https://doi.org/10.1016/j.addr.2015.03.001.

Moodley, J., Cairncross, L., Naiker, T., Momberg, M., 2016. Understanding pathways to breast cancer diagnosis among women in the Western Cape Province, South Africa: A qualitative study. BMJ Open 6. https://doi.org/10.1136/bmjopen-2015-009905.

Moriya, Y., Nakamura, T., Horinouchi, M., Sakaeda, T., Tamura, T., Aoyama, N., Shirakawa, T., Gotoh, A., Fujimoto, S., Matsuo, M., Kasuga, M., Okumura, K., 2002. Effects of polymorphisms of MDR1, MRP1, and MRP2 genes on their mRNA expression levels in duodenal enterocytes of healthy Japanese subjects. Biol. Pharm. Bull. 25, 1356–1359. https://doi.org/10.1248/bpb.25.1356.

Mourskaia, A.A., Amir, E., Dong, Z., Tiedemann, K., Cory, S., Omeroglu, A., Bertos, N., Ouellet, V., Clemons, M., Scheffer, G.L., Park, M., Hallett, M., Komarova, S. V., Siegel, P.M., 2012. ABCC5 supports osteoclast formation and promotes breast cancer metastasis to bone. Breast Cancer Res. 14, 1–16. https://doi.org/10.1186/bcr3361.

Murray, J., Valli, E., Yu, D.M.T., Truong, A.M., Gifford, A.J., Eden, G.L., Gamble, L.D., Hanssen, K.M., Flemming, C.L., Tan, A., Tivnan, A., Allan, S., Saletta, F., Cheung, L., Ruhle, M., Schuetz, J.D., Henderson, M.J., Byrne, J.A., Norris, M.D., Haber, M., Fletcher, J.I., 2017. Suppression of the ATP-binding cassette transporter ABCC4 impairs neuroblastoma tumour growth and sensitises to irinotecan in vivo. Eur. J. Cancer 83, 132–141. https://doi.org/10.1016/j.ejca.2017.06.024.

Nakanishi, T., 2007. Drug transporters as targets for cancer chemotherapy. Cancer Genomics and Proteomics 4, 241–254.

Nalejska, E., Mączyńska, E., Lewandowska, M.A., 2014. Prognostic and predictive

biomarkers: Tools in personalized oncology. Mol. Diagnosis Ther. 18, 273–284. https://doi.org/10.1007/s40291-013-0077-9.

NCR, 2017. National Cancer Registry - South Africa Facts Sheets [WWW Document].

Oeffinger, K., Fontham, E., Etzioni, R., Herzig, A., Michaelson, J., Shih, Y.-C.T., Walter, L., Church, T., Flowers, C., LaMonte, S., Wolf, A., DeSantis, C., Lortet-Tielent, Smith, R., Brawley, O., Wender, R., 2015. Guideline Update from the American Cancer Society. Jama 314, 1599–1614. https://doi.org/10.1001/jama.2015.12783.Breast.

Oguri, T., Ozasa, H., Uemura, T., Bessho, Y., Miyazaki, M., Maeno, K., Maeda, H., Sato, S., Ueda, R., 2008. MRP7/ABCC10 expression is a predictive biomarker for the resistance to paclitaxel in non-small cell lung cancer. Mol. Cancer Ther. 7, 1150–1155. https://doi.org/10.1158/1535-7163.MCT-07-2088.

Ohashi-Kobayashi, A., Ohashi, K., Du, W. Bin, Omote, H., Nakamoto, R., Al-Shawi, M., Maeda, M., 2006. Examination of drug resistance activity of human TAP-like (ABCB9) expressed in yeast. Biochem. Biophys. Res. Commun. 343, 597–601. https://doi.org/10.1016/j.bbrc.2006.03.002.

organization WHO, 2019. World Health Organization (WHO).

Ota, I., Sakurai, A., Toyoda, Y., Morita, S., Sasaki, T., Chishima, T., Yamakado, M., Kawai, Y., Ishidao, T., Lezhava, A., Yoshiura, K.I., Togo, S., Hayashizaki, Y., Ishikawa, T., Ishikawa, T., Endo, I., Shimada, H., 2010. Association between breast cancer risk and the wild-type allele of human ABC transporter ABCC11. Anticancer Res. 30, 5189–5194. https://doi.org/30/12/5189.

Park, S., Shimizu, C., Shimoyama, T., Takeda, M., Ando, M., Kohno, T., Katsumata, N., Kang, Y.K., Nishio, K., Fujiwara, Y., 2006. Gene expression profiling of ATP-binding cassette (ABC) transporters as a predictor of the pathologic response to neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res. Treat. 99, 9–17. https://doi.org/10.1007/s10549-006-9175-2.

Pérez-Herrero, E., Fernández-Medarde, A., 2015. Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy. Eur. J. Pharm. Biopharm. 93, 52–79. https://doi.org/10.1016/j.ejpb.2015.03.018.

Petro-Nustas, W., Tsangari, H., Phellas, C., Constantinou, C., 2013. Health Beliefs and Practice of Breast Self-Examination Among Young Cypriot Women. J. Transcult. Nurs. 24, 180–188. https://doi.org/10.1177/1043659612472201.

Petrucelli, N., Daly, M.B., Pal, T., 2019. BRCA1- and BRCA2 -Associated Hereditary Breast and Ovarian Cancer Summary Genetic counseling Suggestive Findings 1–37.

Pietras, K., Östman, A., 2010. Hallmarks of cancer: Interactions with the tumor stroma. Exp. Cell Res. 316, 1324–1331. https://doi.org/10.1016/j.yexcr.2010.02.045.

Prassas, I., Chrystoja, C.C., Makawita, S., Diamandis, E.P., 2012. Bioinformatic identification of proteins with tissue-specific expression for biomarker discovery. BMC Med. 10, 39. https://doi.org/10.1186/1741-7015-10-39.

Press, M.F., Sauter, G., Buyse, M., Fourmanoir, H., Quinaux, E., Tsao-Wei, D.D., Eiermann, W., Robert, N., Pienkowski, T., Crown, J., Martin, M., Valero, V., Mackey, J.R., Bee, V., Ma, Y., Villalobos, I., Campeau, A., Mirlacher, M., Lindsay, M.A., Slamon, D.J., 2016b. HER2 gene amplification testing by fluorescent in situ hybridization (FISH): Comparison of the ASCO-College of American Pathologists guidelines with FISH scores used for enrollment in Breast Cancer International Research Group clinical trials. J. Clin. Oncol. 34, 3518–3528. https://doi.org/10.1200/JCO.2016.66.6693.

Robey, R.W., Pluchino, K.M., Hall, M.D., Fojo, A.T., Bates, S.E., Gottesman, M.M., 2018. Revisiting the role of ABC transporters in multidrug-resistant cancer. Nat. Rev. Cancer 18, 452–464. https://doi.org/10.1038/s41568-018-0005-8.

Roth, M., Obaidat, A., Hagenbuch, B., 2012. OATPs, OATs and OCTs: The organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. Br. J. Pharmacol. 165, 1260–1287. https://doi.org/10.1111/j.1476-5381.2011.01724.

Ruddy, K.J., Desantis, S.D., Gelman, R.S., Wu, A.H.B., Punglia, R.S., Mayer, E.L., Tolaney, S.M., Winer, E.P., Partridge, A.H., Burstein, H.J., 2013. Personalized medicine in breast cancer: Tamoxifen, endoxifen, and CYP2D6 in clinical practice. Breast Cancer Res. Treat. 141, 421–427. https://doi.org/10.1007/s10549-013-2700-1.

Russo, J., Hu, Y.F., Yang, X., Russo, I.H., 2000. Developmental, cellular, and molecular basis of human breast cancer. J. Natl. Cancer Inst. Monogr. 2000, 17–37. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024241.

Sharan, S.K., J, M.M., Albrecht, U., J, D.L., Regelt, E., Dinh, C., Sands, A., Eichele, G., J, P.H., Bradley, A., 1995. hypersensitivity mediated by Rad51 386124, 6408–6412.

Sharom, F.J., 2008. ABC multidrug transporters: Structure, function and role in chemoresistance. Pharmacogenomics 9, 105–127. https://doi.org/10.2217/14622416.9.1.10.

Shu, Y., Leabman, M.K., Feng, B., Mangravite, L.M., Huang, C.C., Stryke, D., Kawamoto, M., Johns, S.J., DeYoung, J., Carlson, E., Ferrin, T.E., Herskowitz, I., Giacomini, K.M., Benet, L.Z., Brett, C.M., Burchard, E.G., Castro, R., De la Cruz, M., Edwards, R.H., Gitschier, J., Glatt, C.E., Ho, C., Kroetz, D.L., Lin, E.T., Reus, V.I., Sadee, W., Salazar, M., Schaefer, C., Sheiner, L.B., Tran, C., Urban, T.J., Vulpe, C., Wright, E.M., 2003. Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. Proc. Natl. Acad. Sci. U. S. A. 100, 5902–5907. https://doi.org/10.1073/pnas.0730858100.

Shukla, S., Ohnuma, S., V. Ambudkar, S., 2011. Improving Cancer Chemotherapy with Modulators of ABC Drug Transporters. Curr. Drug Targets 12, 621–630. https://doi.org/10.2174/138945011795378540.

Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S.D., Willis, J., Dawson, D., Willson, J.K.V., Gazdar, A.F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B.H., Bachman, K.E., Papadopoulos, N., Vogelstein, B., Kinzler, K.W., Velculescu, V.E., 2006b. The consensus coding sequences of human breast and colorectal cancers. Science (80-.). 314, 268–274. https://doi.org/10.1126/science.1133427.

Slaoui, M., Razine, R., Ibrahimi, A., Attaleb, M., El Mzibri, M., Amrani, M., 2014. Breast cancer in Morocco: A literature review. Asian Pacific J. Cancer Prev. 15, 1067–1074. https://doi.org/10.7314/APJCP.2014.15.3.1067.

Soleas, G.J., Diamandis, E.P., Goldberg, D.M., 1997. Wine as a biological fluid: History, production, and role in disease prevention. J. Clin. Lab. Anal. 11, 287–313. https://doi.org/10.1002/(SICI)1098-2825(1997)11:5<287::AID-JCLA6>3.0.CO;2-4.

Sosonkina, N., Nakashima, M., Ohta, T., Niikawa, N., Starenki, D., 2011. Down-regulation of

ABCC11 protein (MRP8) in human breast cancer. Exp. Oncol. 33, 42–46.

Stefansson, O.A., Jonasson, J.G., Johannsson, O.T., Olafsdottir, K., Steinarsdottir, M., Valgeirsdottir, S., Eyfjord, J.E., 2009. Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. Breast Cancer Res. 11, 1–14. https://doi.org/10.1186/bcr2334.

Stenson, P.D., Ball, E. V., Mort, M., Phillips, A.D., Shaw, K., Cooper, D.N., 2012. The human gene mutation database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. Curr. Protoc. Bioinforma. Chapter 1, Unit1.13. https://doi.org/10.1002/0471250953.bi0113s39.

Sun, J., Xu, J., Li, Y., Shi, T., 2005a. Analysis and application of large-scale protein-protein interaction data sets. Chinese Sci. Bull. 50, 2267–2272. https://doi.org/10.1360/982005-327.

Sun, Y.L., Patel, A., Kumar, P., Chen, Z.S., 2012. Role of ABC transporters in cancer chemotherapy. Chin. J. Cancer 31, 51–57. https://doi.org/10.5732/cjc.011.10466.

Suresh, P.K., Varadharaj, V.A., Mathiyazhagan, J., 2015. These transporters (highly conserved throughout. Int J Drug Dev Res 7, 127–136.

Szakács, G., Váradi, A., Özvegy-Laczka, C., Sarkadi, B., 2008. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). Drug Discov. Today 13, 379–393. https://doi.org/10.1016/j.drudis.2007.12.010.

Tabarestani, S., Ghaderian, S.M.H., Rezvani, H., Mirfakhraie, R., 2014a. Expression profiling of breast cancer patients treated with tamoxifen: Prognostic or predictive significance. Med. Oncol. 31. https://doi.org/10.1007/s12032-014-0896-5.

Tian, Y., Tian, X., Han, X., Chen, Y., Song, C.Y., Jiang, W.J., Tian, D.L., 2016. ABCE1 plays an essential role in lung cancer progression and metastasis. Tumor Biol. 37, 8375–8382. https://doi.org/10.1007/s13277-015-4713-3.

Toyoda, Y., Ishikawa, T., 2011. Pharmacogenomics of Human ABC Transporter ABCC11 (MRP8): Potential Risk of Breast Cancer and Chemotherapy Failure. Anticancer. Agents Med. Chem. 10, 617–624. https://doi.org/10.2174/187152010794473975.

Tyrer, J., Duffy, S.W., Cuzick, J., 2004. A breast cancer prediction model incorporating

familial and personal risk factors. Stat. Med. 23, 1111–1130. https://doi.org/10.1002/sim.1668.

Useful Molecular Modelling and Drug Design Softwares and Databases, 2012. https://doi.org/10.1021/ci3001277.ChEMBL.

Van De Vijver, M.J., 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., Rutgers, E.J.T., Peterse, J.L., Halfwerk, H., De Groot, R., Pronk, A., Floore, A.N., Glas, A.M., Van't Veer, L.J., 2009. Validation of 70-gene prognosis signature in node-negative breast cancer. Breast Cancer Res. Treat. 117, 483–495. https://doi.org/10.1007/s10549-008-0191-2.

VictorDevadoss, A., Janet Sheeba, J., Albert William, M., 2013. Finding the Peak Age of an Indian Woman Victim of Breast Cancer using CETD Matrix. Int. J. Comput. Appl. 78, 26–32. https://doi.org/10.5120/13527-1257.

Waks, A.G., Winer, E.P., 2019. Breast Cancer Treatment: A Review. JAMA - J. Am. Med. Assoc. 321, 288–300. https://doi.org/10.1001/jama.2018.19323.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., Kerr, D., 2009. Genetic prognostic and predictive markers in colorectal cancer. Nat. Rev. Cancer 9, 489–499. https://doi.org/10.1038/nrc2645.

Wang, J., Huo, K., Ma, L., Tang, L., Li, D., Huang, X., Yuan, Y., Li, C., Wang, W., Guan, W., Chen, H., Jin, C., Wei, J., Zhang, W., Yang, Y., Liu, Q., Zhou, Y., Zhang, C., Wu, Z., Xu, W., Zhang, Y., Liu, T., Yu, D., Zhang, Y., Chen, L., Zhu, D., Zhong, X., Kang, L., Gan, X., Yu, X., Ma, Q., Yan, J., Zhou, L., Liu, Z., Zhu, Y., Zhou, T., He, F., Yang, X., 2011. Toward an understanding of the protein interaction network of the human liver. Mol. Syst. Biol. 7, 1–10. https://doi.org/10.1038/msb.2011.67.

Weigel, M.T., Dowsett, M., 2010. Current and emerging biomarkers in breast cancer: Prognosis and prediction. Endocr. Relat. Cancer 17, 245–262. https://doi.org/10.1677/ERC-10-0136.

Whirl-Carrillo, M., McDonagh, E.M., Hebert, J.M., Gong, L., Sangkuhl, K., Thorn, C.F., Altman, R.B., Klein, T.E., 2012. Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 92, 414–417. https://doi.org/10.1038/clpt.2012.96.

Wijnholds, J., Mol, C.A.A.M., Van Deemter, L., De Haas, M., Scheffer, G.L., Baas, F., Beijnen, J.H., Scheper, R.J., Hatse, S., De Clercq, E., Balzarini, J., Borst, P., 2000. Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. Proc. Natl. Acad. Sci. U. S. A. 97, 7476–7481. https://doi.org/10.1073/pnas.120159197.

Wojtal, K.A., Eloranta, J.J., Hruz, P., Gutmann, H., Drewe, J., Staumann, A., Beglinger, C., Fried, M., Kullak-Ublick, G.A., Vavricka, S.R., 2009. Changes in mRNA expression levels of solute carrier transporters in inflammatory bowel disease patients. Drug Metab. Dispos. 37, 1871–1877. https://doi.org/10.1124/dmd.109.027367.

Yamada, A., Ishikawa, T., Ota, I., Kimura, M., Shimizu, D., Tanabe, M., Chishima, T., Sasaki, T., Ichikawa, Y., Morita, S., Yoshiura, K.I., Takabe, K., Endo, I., 2013. High expression of ATP-binding cassette transporter ABCC11 in breast tumors is associated with aggressive subtypes and low disease-free survival. Breast Cancer Res. Treat. 137, 773–782. https://doi.org/10.1007/s10549-012-2398-5.

Yang, Y., Adelstein, S.J., Kassis, A.I., 2009. Target discovery from data mining approaches. Drug Discov. Today 14, 147–154. https://doi.org/10.1016/j.drudis.2008.12.005.

Yang, Y., Dong, X., Xie, B., Ding, N., Chen, J., Li, Y., Zhang, Q., Qu, H., Fang, X., 2015. Databases and web tools for cancer genomics study. Genomics, Proteomics Bioinforma. 13, 46–50. https://doi.org/10.1016/j.gpb.2015.01.005.

Yasui, K., Mihara, S., Zhao, C., Okamoto, H., Saito-Ohara, F., Tomida, A., Funato, T., Yokomizo, A., Naito, S., Imoto, I., Tsuruo, T., Inazawa, J., 2004. Alteration in Copy Numbers of Genes as a Mechanism for Acquired Drug Resistance. Cancer Res. 64, 1403–1410. https://doi.org/10.1158/0008-5472.CAN-3263-2.

Yoshida, M., Suzuki, T., Komiya, T., Hatashita, E., Nishio, K., Kazuhiko, N., Fukuoka, M., 2001. Induction of MRP5 and SMRP mRNA by adriamycin exposure and its overexpression in human lung cancer cells resistant to Adriamycin. Int. J. Cancer 94, 432–437. https://doi.org/10.1002/ijc.1490.

Yu, X., Lin, J., Zack, D.J., Qian, J., 2006. Computational analysis of tissue-specific combinatorial gene regulation: Predicting interaction between transcription factors in human tissues. Nucleic Acids Res. 34, 4925–4936. https://doi.org/10.1093/nar/gkl595.

Zhang, H., Lian, Z., Sun, G., Liu, R., Xu, Y., 2018. Loss of miR-516a-3p mediates upregulation of ABCC5 in prostate cancer and drives its progression. Onco. Targets. Ther. 11, 3853–3867. https://doi.org/10.2147/OTT.S167463.

Zhang, Y.K., Wang, Y.J., Gupta, P., Chen, Z.S., 2015. Multidrug Resistance Proteins (MRPs) and Cancer Therapy. AAPS J. 17, 802–812. https://doi.org/10.1208/s12248-015-9757-1.

Zhang, Z., Wang, J., Shen, B., Peng, C., Zheng, M., 2012. The ABCC4 gene is a promising target for pancreatic cancer therapy. Gene 491, 194–199. https://doi.org/10.1016/j.gene.2011.09.029.

Zhao, H., Huang, Y., Shi, J., Dai, Y., Wu, L., Zhou, H., 2018. ABCC10 plays a significant role in the transport of gefitinib and contributes to acquired resistance to gefitinib in NSCLC. Front. Pharmacol. 9, 1–13. https://doi.org/10.3389/fphar.2018.01312.

Zhao, L., Yang, X., Khan, A., Kandil, D., 2014. Diagnostic role of immunohistochemistry in the evaluation of breast pathology specimens. Arch. Pathol. Lab. Med. 138, 16–24. https://doi.org/10.5858/arpa.2012-0440-RA.

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J.R., Mohammed, S., 2013. Toward a comprehensive characterization of a human cancer cell phosphoproteome. J. Proteome Res. 12, 260–271. https://doi.org/10.1021/pr300630k.

Zhou, Y., 2013. Noninvasive treatment of breast cancer using high-intensity focused ultrasound. J. Med. Imaging Heal. Informatics 3, 141–156. https://doi.org/10.1166/jmihi.2013.1156.

Zhou, Y., Zhang, G.Q., Wei, Y.H., Zhang, J.P., Zhang, G.R., Ren, J.X., Duan, H.G., Rao, Z., Wu, X.A., 2013. The impact of drug transporters on adverse drug reaction. Eur. J. Drug Metab. Pharmacokinet. 38, 77–85. https://doi.org/10.1007/s13318-013-0117-1.

Appendix A

Chapter Two: Supplementary Information.

TiGER and HPA expression profiles of the of ABCB2, ABCB9, ABCB10, ABCC1, ABCC4, ABCC12 and ABCD1 genes.

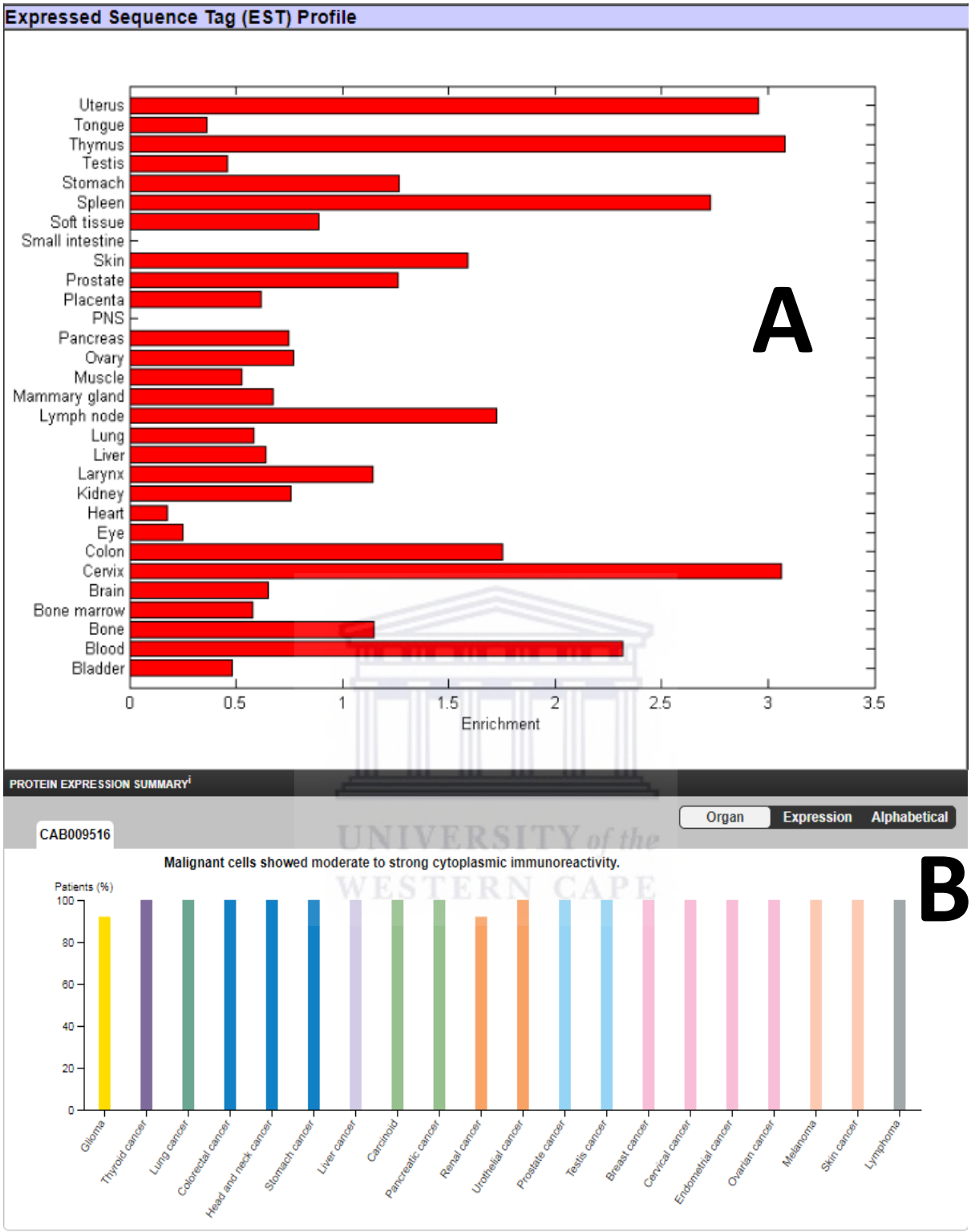

Figure 2.4: Expression profile for ABCB2 from (A) TiGER and (B) HPA.

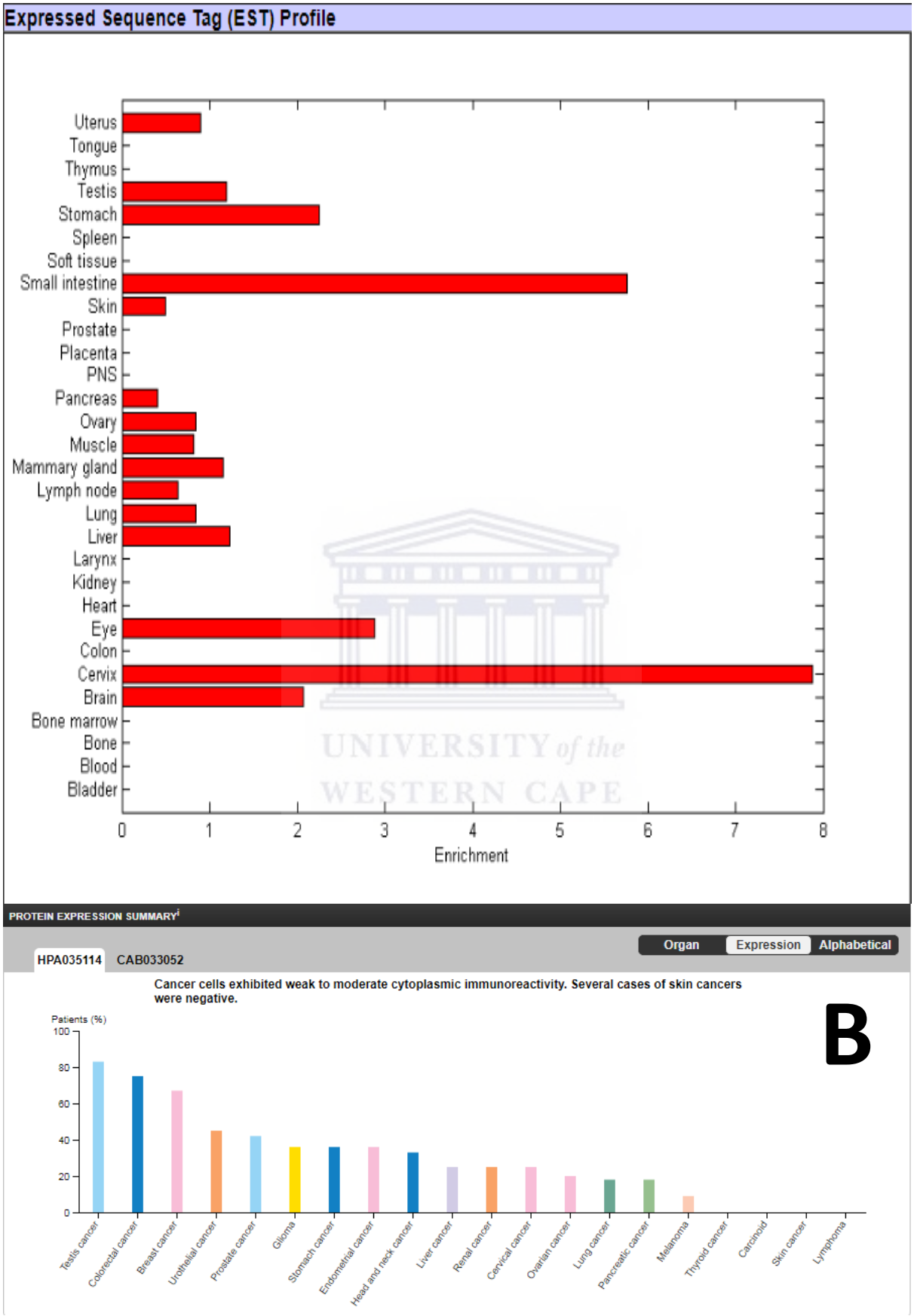

Figure 2.5: Expression profile for ABCB9 from (A) TiGER and (B) HPA.

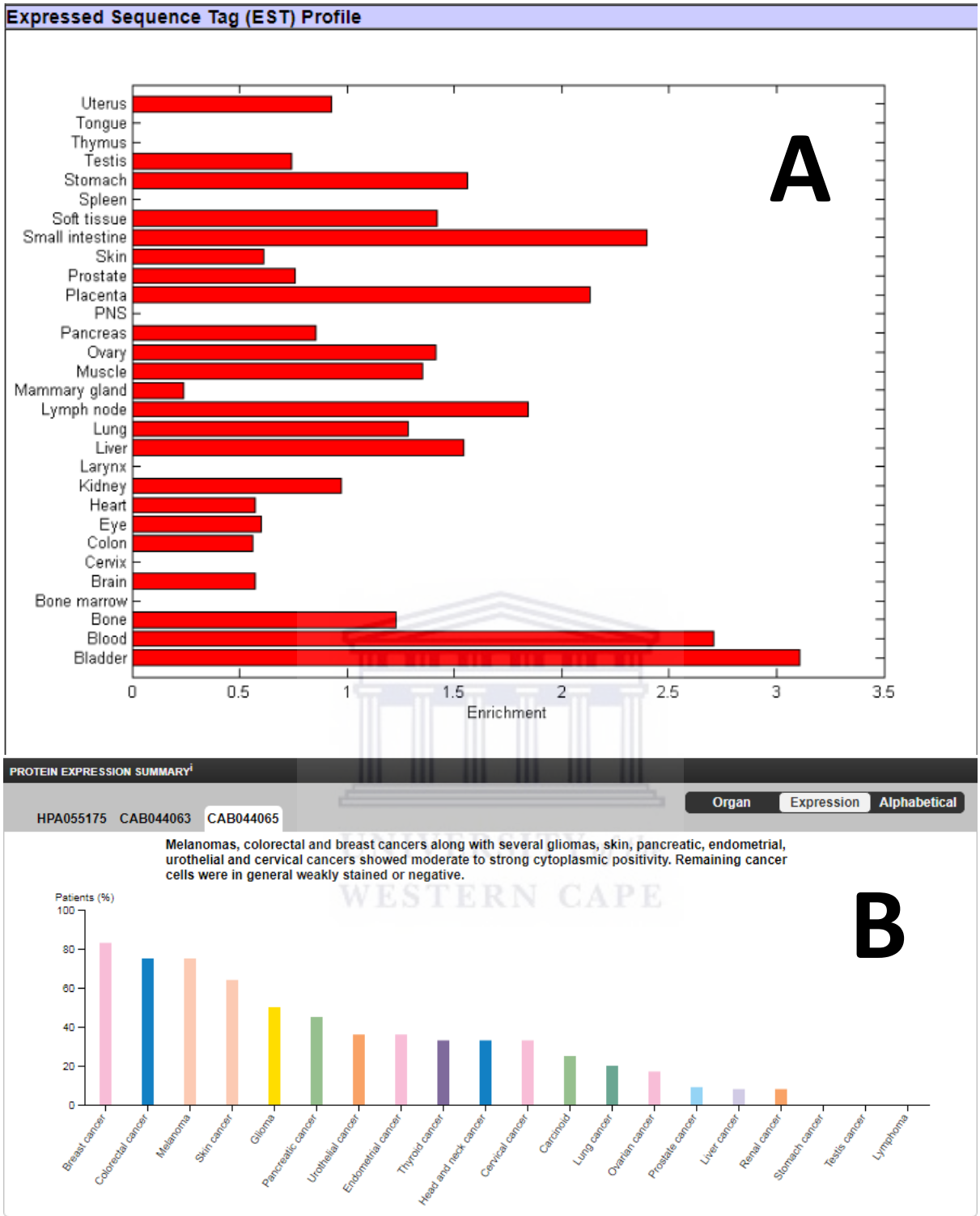

Figure 2.6: Expression profile for ABCB10 from (A) TiGER and (B) HPA.

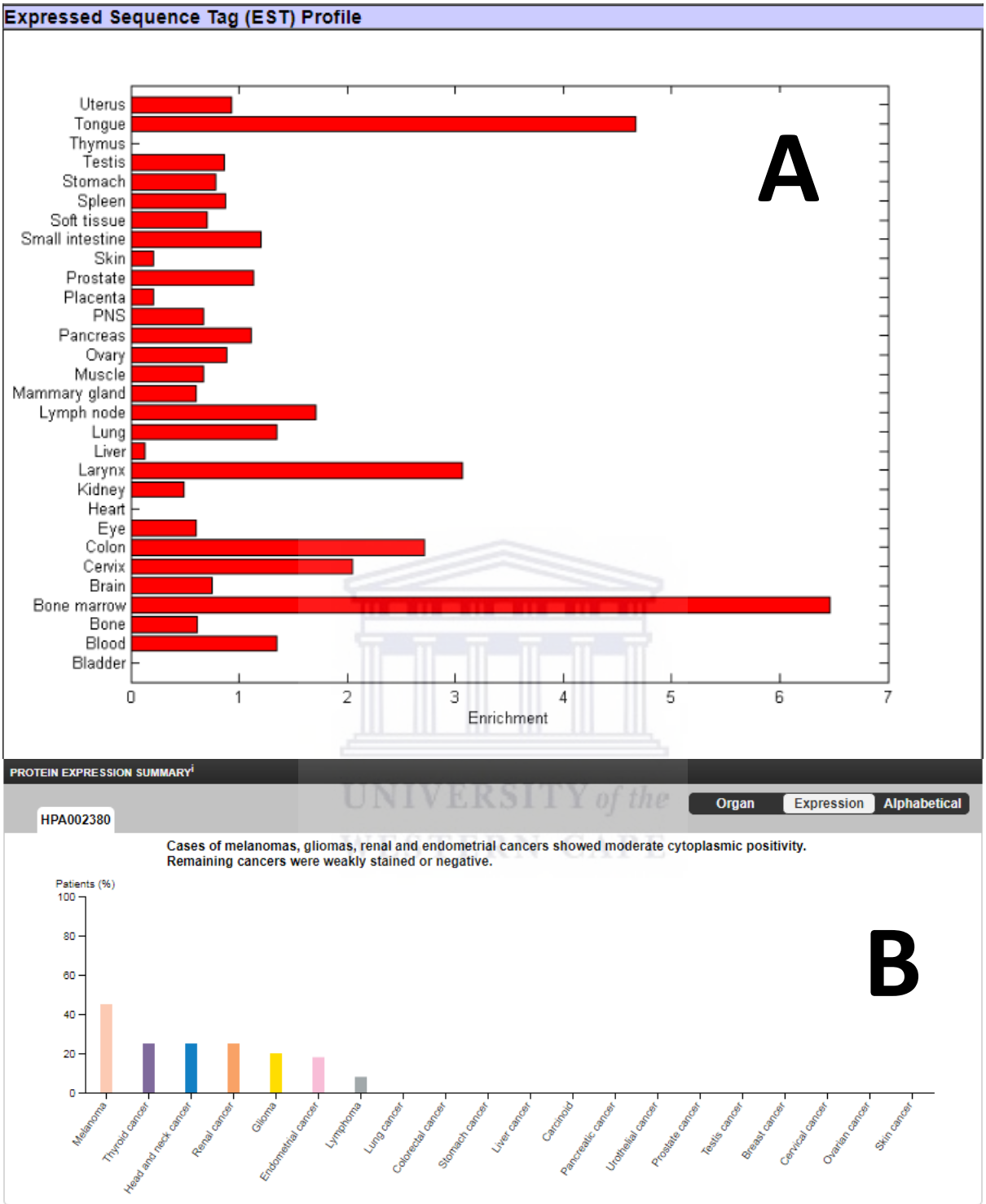

Figure 2.7: Expression profile for ABCC1 from (A) TiGER and (B) HPA.

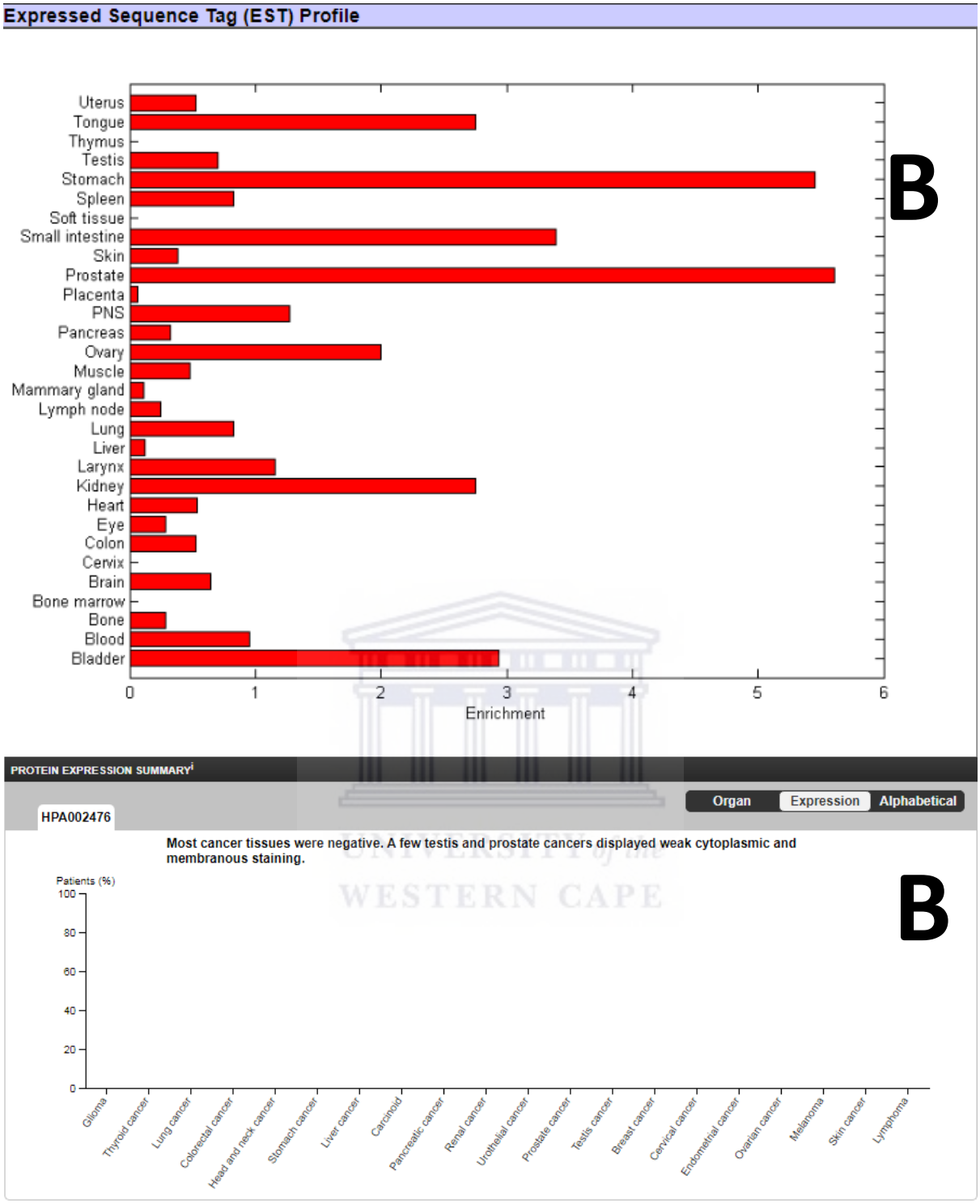

Figure 2.8: Expression profile for ABCC4 from (A) TiGER and (B) HPA.

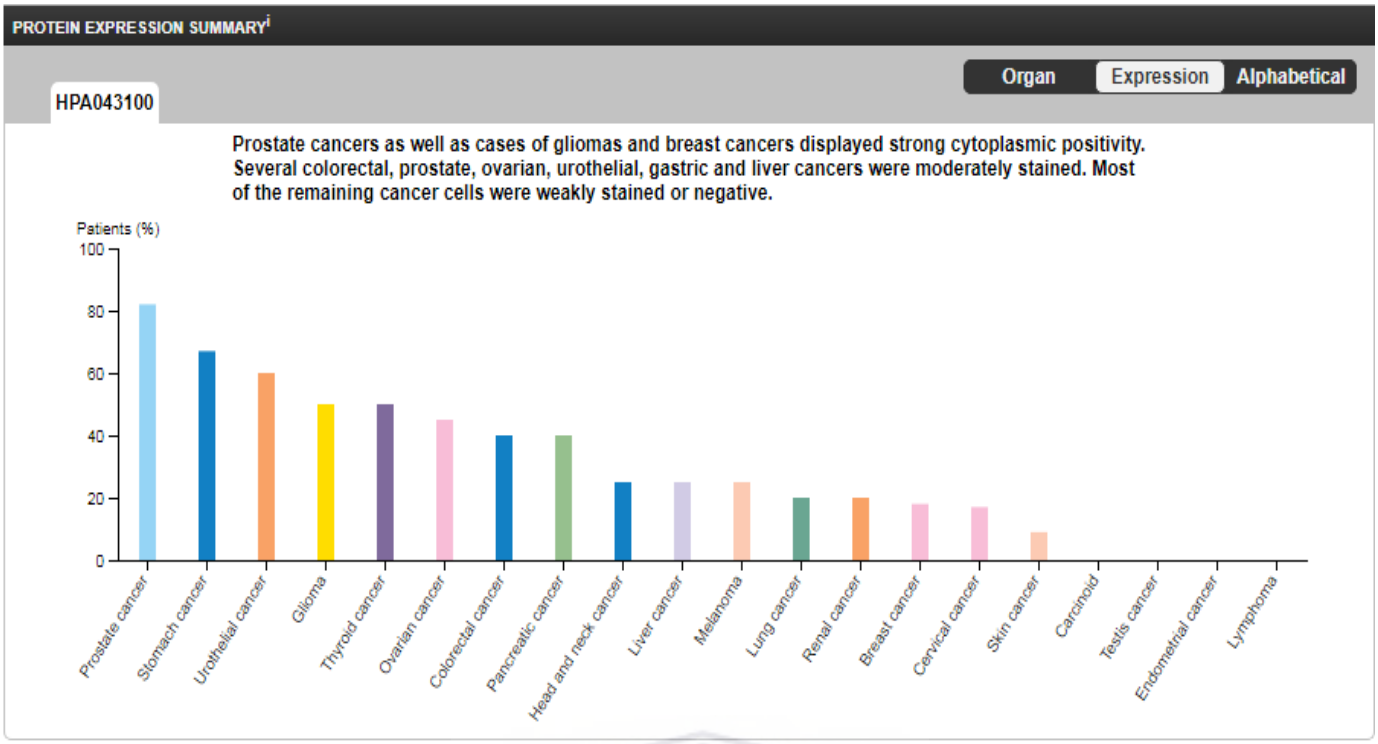

Figure 2.9: Expression profile for ABCC12 from HPA. However, ABCC12 obtained no result on the TiGER database.

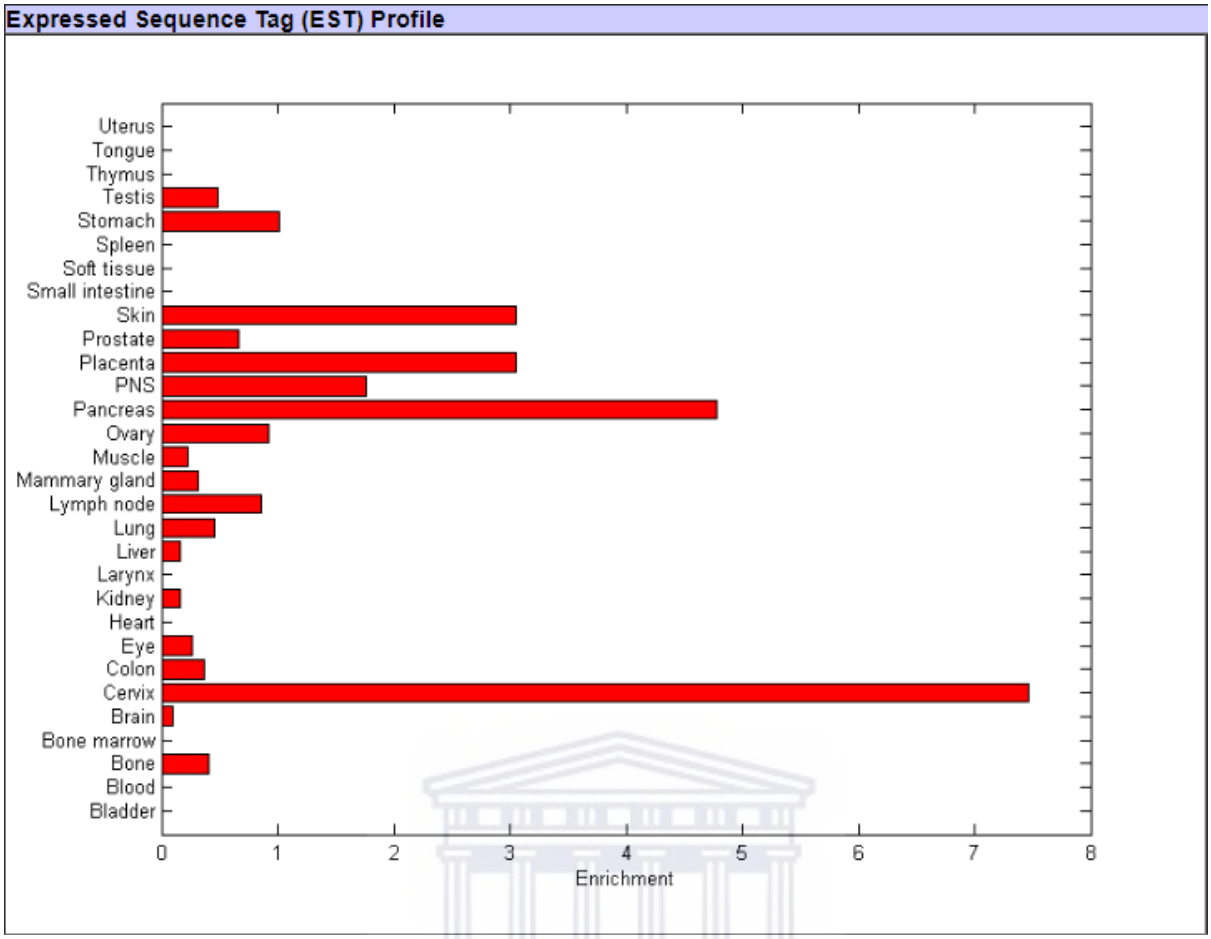

Figure 2.10: Expression profile for ABCD1 from TiGER database. However, ABCD1 obtained no result on the HPA.

Appendix **B**

Chapter three: Supplementary Information.

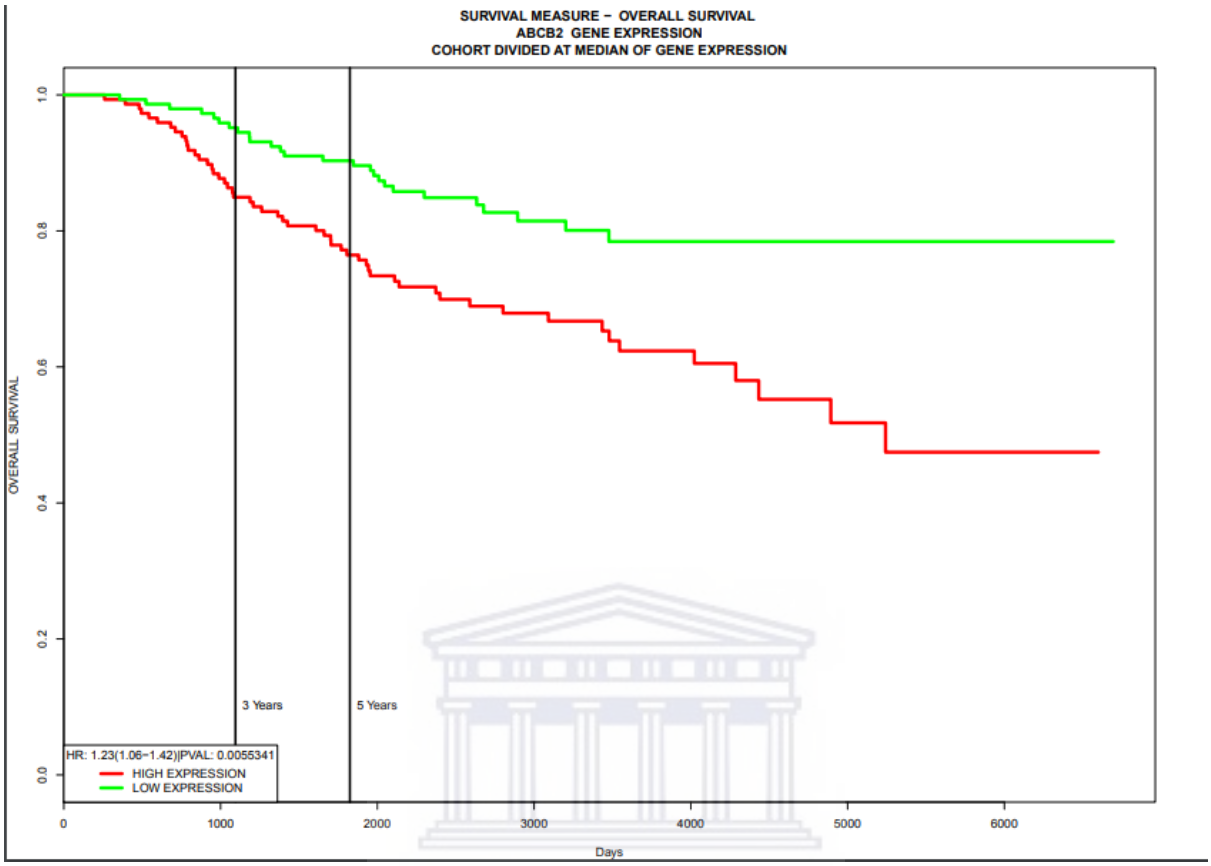

Figure 3.5: Overall survival analysis profile for ABCB2 gene using PROGGENE showing high (red line) and low (green line) levels of expression at different time intervals.

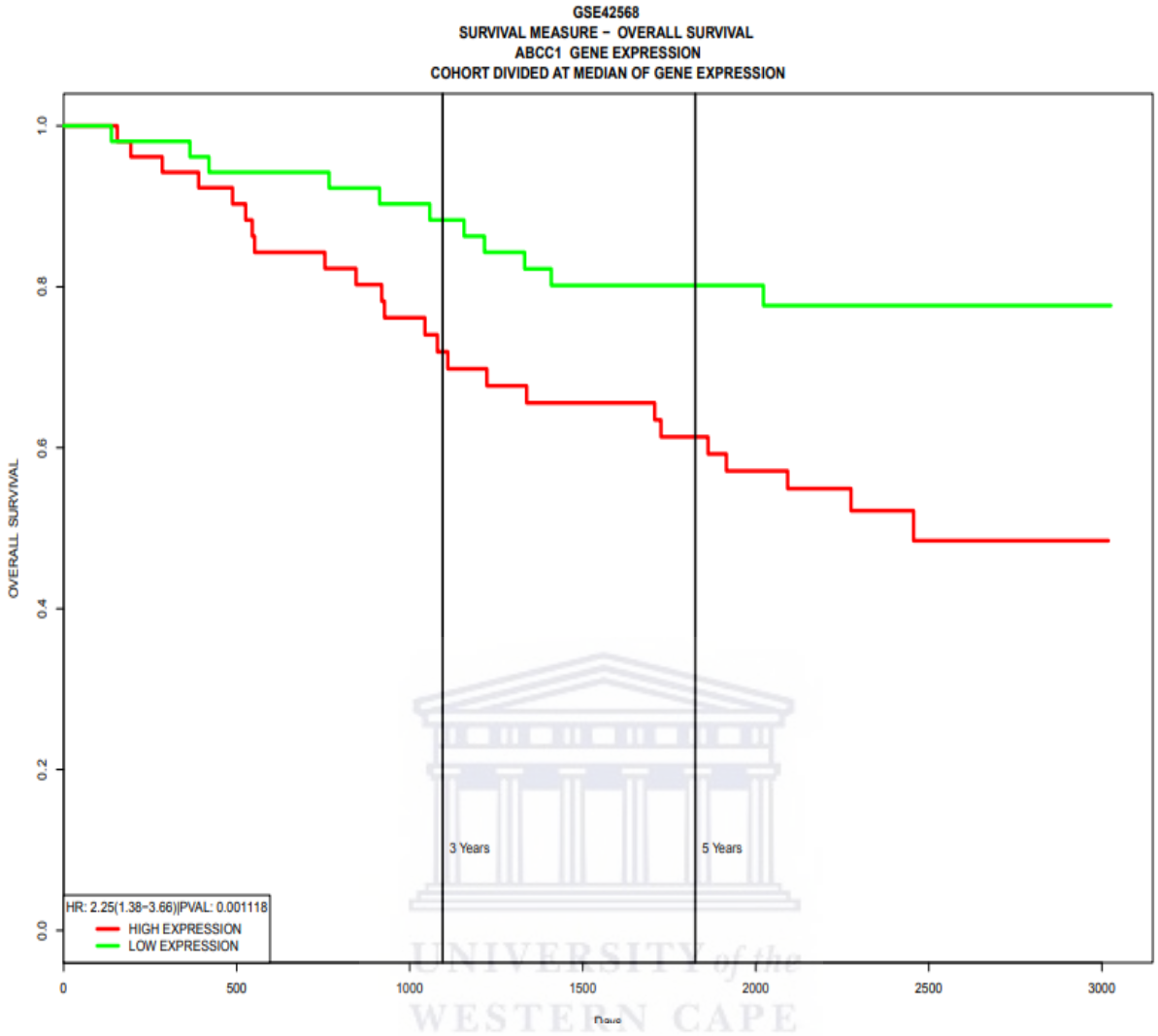

Figure 3.6: Overall survival analysis profile for ABCC1 gene using PROGGENE showing high (red line) and low (green line) levels of expression at different time intervals.

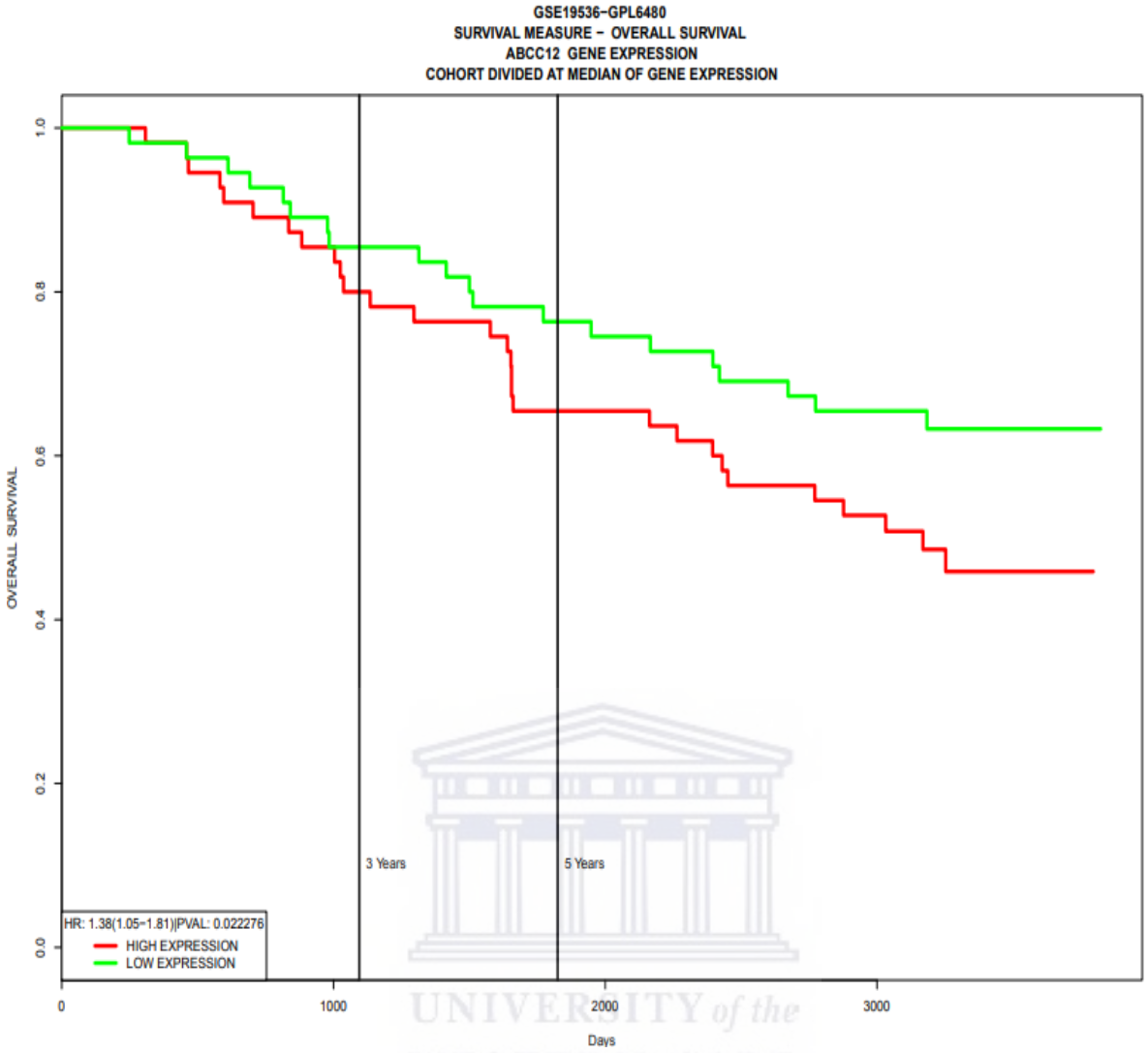

Figure 3.7: Overall survival analysis profile for ABCC12 gene using PROGGENE showing high (red line) and low (green line) levels of expression at different time intervals.

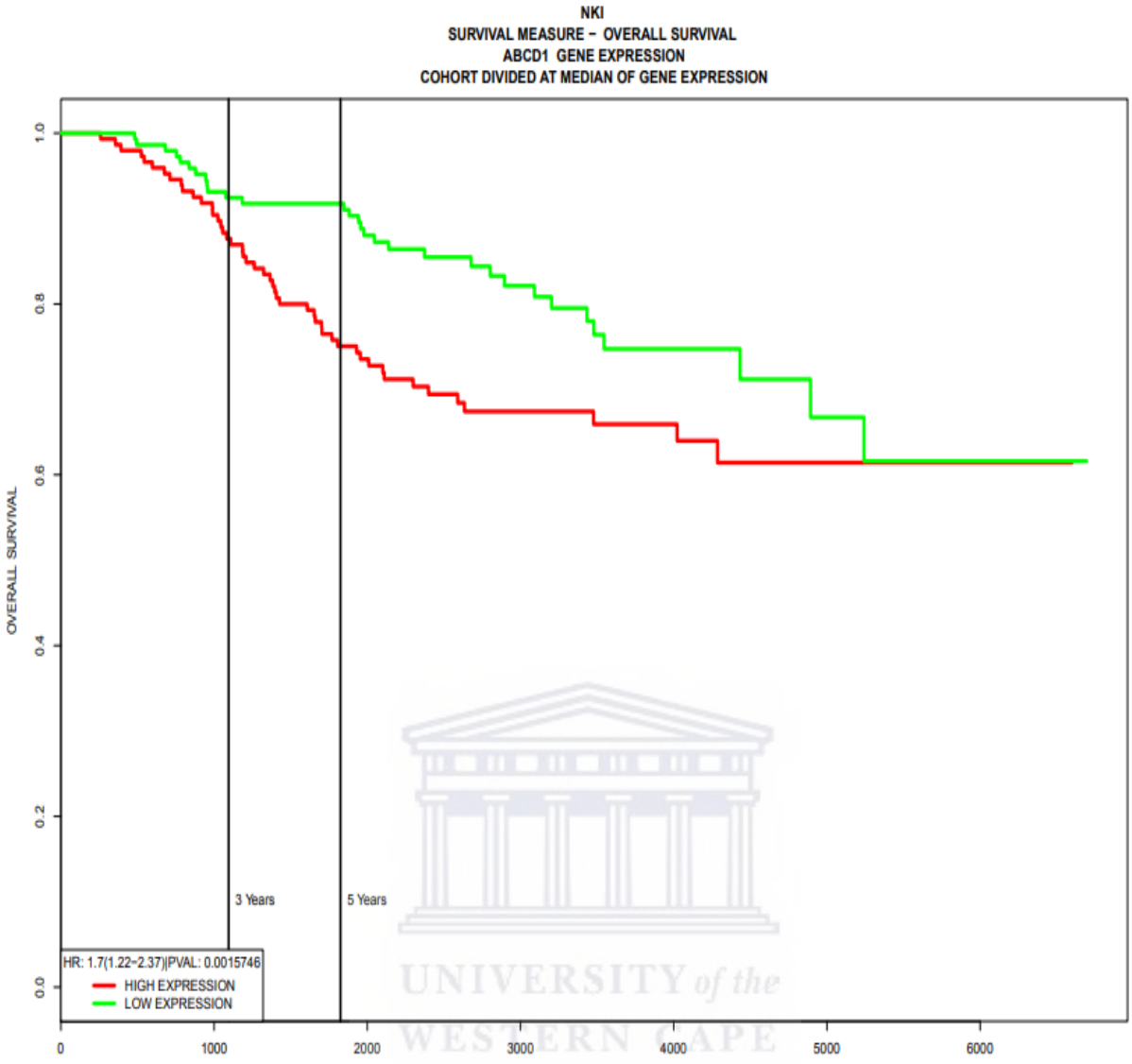

Figure 3.8: Overall survival analysis profile for ABCD1 gene using PROGGENE showing high (red line) and low (green line) levels of expression at different time intervals.

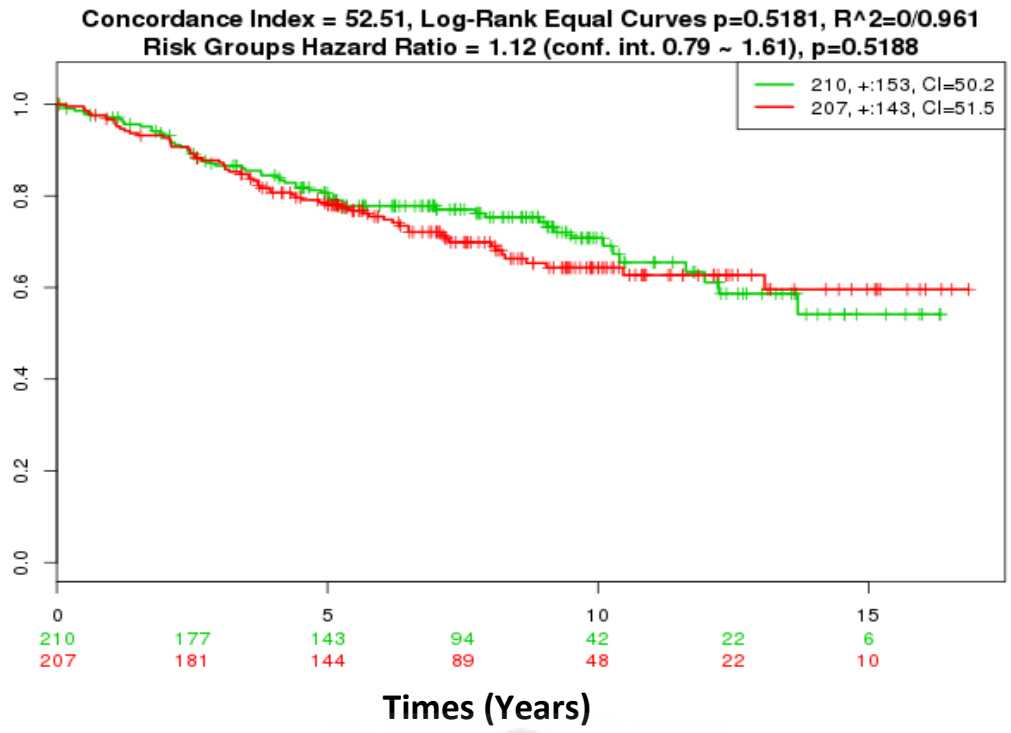

Figure 3.9: Prognostic gene expression analysis of ABCB2 using SurvExpress. The expression analysis is captured in Kaplan Meier plot showing the low (green) and high (red) gene expression in risk groups respectively using the log rank test.

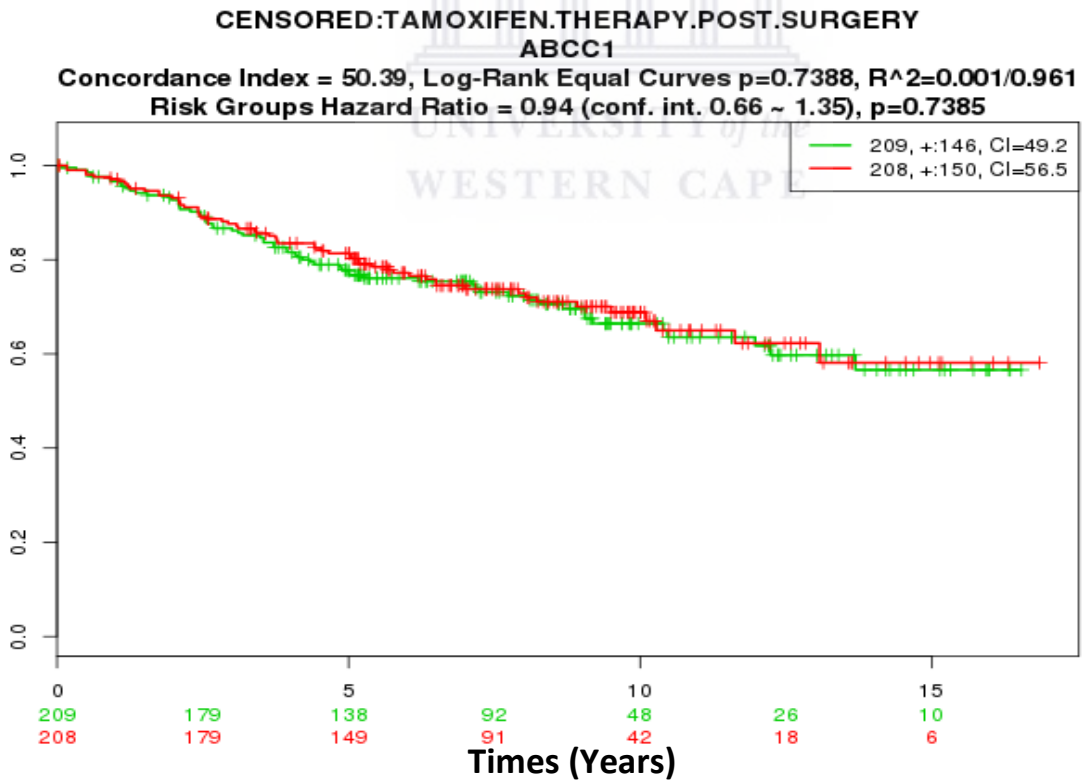

Figure 3.10: Prognostic gene expression analysis of ABCC1 using SurvExpress. The expression analysis is captured in Kaplan Meier plot showing the low (green) and high (red) gene expression in risk groups respectively using the log rank test.

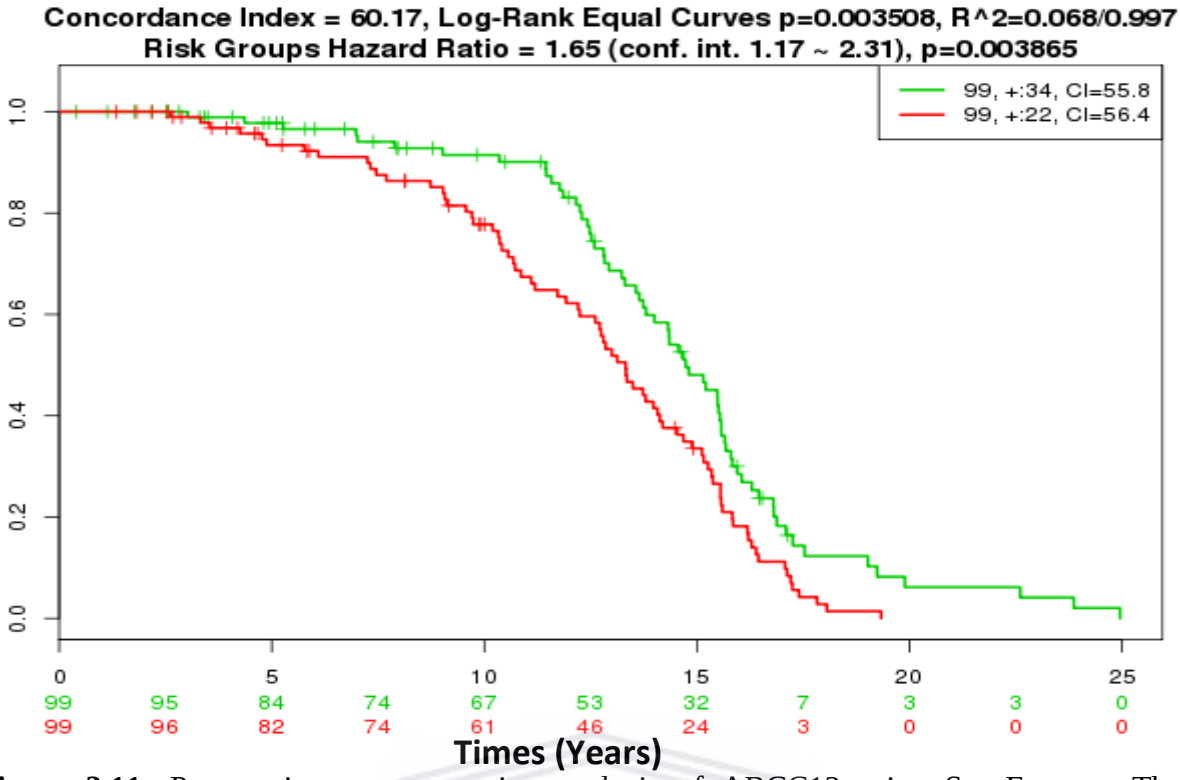

Figure 3.11: Prognostic gene expression analysis of ABCC12 using SurvExpress. The expression analysis is captured in Kaplan Meier plot showing the low (green) and high (red) gene expression in risk groups respectively using the log rank test.

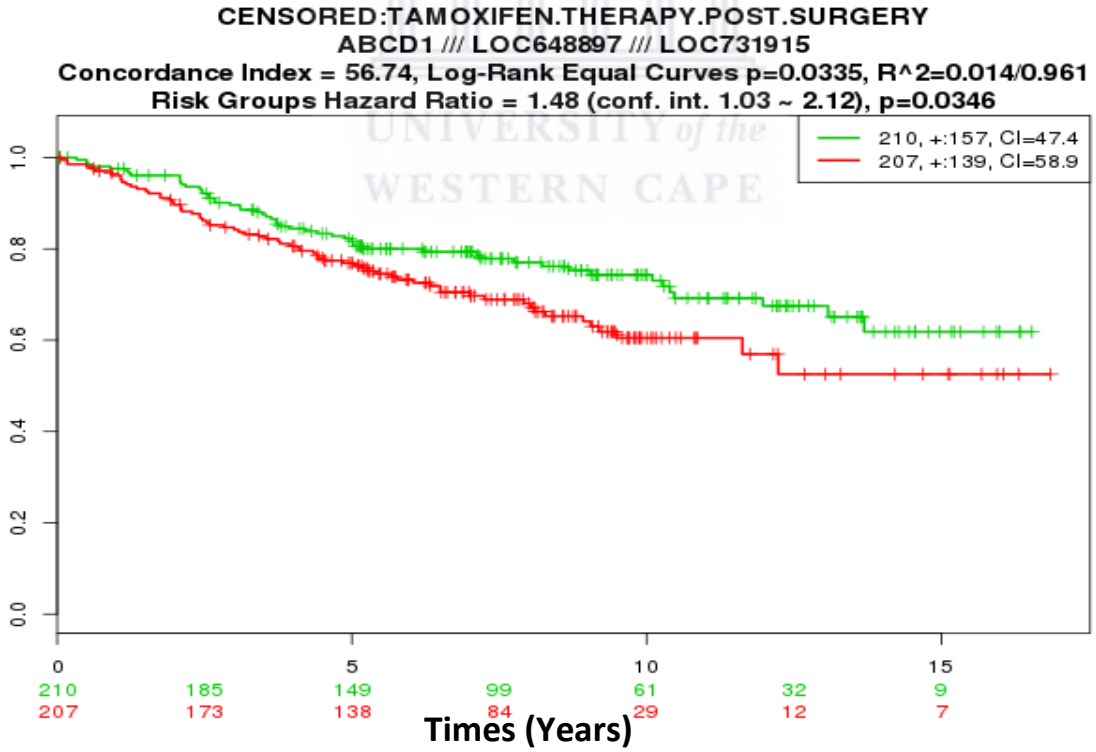

Figure 3.12: Prognostic gene expression analysis of ABCD1 using SurvExpress. The expression analysis is captured in Kaplan Meier plot showing the low (green) and high (red) gene expression in risk groups respectively using the log rank test.