GENOMIC INSTABILITY IN SOUTH AFRICAN BREAST CANCER PATIENTS

by

Bridget Cebisile Langa

Submitted in fulfilment of the requirements for the degree

Magister Scientiae (M.Sc.) in Medical Biosciences

Department of Medical Biosciences
University of the Western Cape
South Africa

Supervisor: Dr Sahar Abdul-Rasool

Co-supervisor: Prof Donavon Hiss

November 2013

DECLARATION

I, **Bridget Cebisile Langa**, hereby declare that the dissertation "**Genomic Instability in South African Breast Cancer Patients**" for the Masters degree in Science (MSc) at the University of the Western Cape hereby submitted by me has not been submitted previously at this or any other university, and that it is my own work in design and in execution, and that all materials contained herein have been duly acknowledged by complete references.

Bridget Cebisile Langa **:** .....

Date Signed **:** .....

DEDICATION

This dissertation is dedicated to my parents, Bhekokwakhe Langa and Nomakhuze Langa who have encouraged me throughout the years as I persue my academic endeavours, and for their support and love. To my siblings, Nokulunga Madikizela, Hloniphile Langa, Sandisiwe Langa, Velenkosini Langa and Thubelihle Langa, who have always believed in me. I thank you all for your powerful prayers, love and support throughout the years of my academics.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Almighty God for granting me the opportunity to pursue and complete my degree.

My sincere gratitude and appreciation goes to my father and mother without whom I would not be who I am today.

I express my sincere gratitude and appreciation to my supervisors Dr Sahar Abdul-Rasool and Prof Donavon Hiss, for their support, supervision, guidance and encouragement.

A word of appreciation goes to Georgetown University, Lombardi Cancer Center (United States of America) for allowing me access to their laboratory and the training I received on the techniques used in this dissertation.

Heartfelt thanks goes to my colleagues and staff, especially Dr L. Mbonile, Ms. N. Haines, Mr Y Mnyamana and everyone in the Department of Medical Biosciences, University of the Western Cape.

A word of appreciation to the Cancer Association of South Africa (CANSA) for their financial support of this work.

ABSTRACT

Genomic instability in South African breast cancer patients

B.C. Langa, Degree of *Magister Scientiae* (MSc) Thesis, Department of Medical Biosciences, University of the Western Cape

Breast cancer (BC) is one of the most common malignancies in women. Death results from treatment failure and metastatic disease. Thousands of lives might be saved if it was possible to detect and eliminate occult metastatic cells before they become clinically evident. Therefore, there is a critical need to identify new markers to improve treatment options for these patients.

Genomic instability is the earliest indication of breast cancer and the use of genomic methodologies is a progress towards early detection and treatment, through the identification of biomarkers that can be translated into novel therapy targets.

The interferon regulatory factor-1*(IRF-1)* gene, localized on chromosome 5q31.1, is believed to act as a tumor suppressor gene in breast cancer. The *IRF-1* was found to be inactivated by single nucleotide polymorphism (SNP) in breast cancer suggesting that the loss of its function might be critical to the development of the disease.

The phosphatidylinositol 3-kinase *(PIK3)* signaling pathway mediates key cellular functions and alterations of genes in this pathway, including *PIK3CA*, serine-threonine protein kinases (*AKT1and AKT2)*, phosphatase and tensin homolog *(PTEN)*, fibroblast growth factor receptor 2 (*FGFR2)* and *ERBB2*, whose expression have been demonstrated to be altered in breast cancer patients. In addition, these genes are linked to treatment resistance.

In this study, we have investigated allelic loss of *IRF-1* gene in primary tumors obtained from patients undergoing mastectomy at Groote Schuur hospital (Cape Town, South Africa). These samples were then further analyzed for the DNA copy number changes of specific genes involved in the *PIK3/AKT* signaling pathway.

Statistical analysis has been performed in order to correlate genomic findings with clinical-histopathological and follow up information from the patients and to establish whether these genes can predict prognosis.

Our data analysis has indicated that 46 cases (45.5%) out of 101 cases were informative for the *IRF-1* dinucleotide marker used for LOH analysis (Figure 3.1). LOH was detected in 23 of these informative cases (23/46; 50%). No statistical significance was found between LOH at the *IRF-1* locus and age (≤50 years or >50 years) (P value = 1.0000) and earlier stage (Stages I and II) (P value= 0.4982) based on Fisher's exact test.

Patients presented a high level of DNA copy number changes in genes involved in the *PIK3/AKT* pathway. The most frequent changes were observed in the *PIK3CA* and *PTEN* genes. *PIK3CA* presented high copy number in 36.8% of the cases. *PTEN* was observed with low copy number in 47.5% of the cases.

This dissertation shows the effectiveness of genomic methodologies as means for the detection of early breast cancer progression in South African women. The *PIK 3/AKT* genes can validate the usefulness of breast cancer therapies.

Keywords: Breast cancer, primary tumors, loss of heterozygosity, genomic instability, the interferon regulator factor 1, the phosphatidyl inositol kinase pathway.

LIST OF ABBREVIATIONS

AKT1	Serine-threonine protein kinases 1
AKT2	Serine-threonine protein kinases 2
BC	Breast Cancer
bp	Base pairs
BMI	Body mass index
CANSA	Cancer Association of South Africa
CAE	Capillary array electrophoresis
ºC	Degrees Celsius
CNVs	Copy number variations
C_t	Crossing threshold
DL	Ductal lavage
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
DSBs	Double-strand breaks
ER	Oestrogen receptor
FFPE	Formalin-fixed, paraffin-embedded
FGFR2	Fibroblast growth factor receptor 2
GDB	Genome Data Base
H&E	Haematoxylin & Eosin
IRF -1	Interferon regulatory factor-1
LOH	Loss of heterozygosity
LM	Laser-assisted microdissection
mRNA	Messenger ribonucleic acid
mTOR	Mammalian target of rapamycin
mM	Micromolar
MgCl2	Magnesium chloride

NLHS	National Laboratory Health Sciences
NAF	Nipple aspirate fluid
PR	Progesterone receptor
PCR	Polymerase chain reaction
PTEN	Phosphatase and tensin homolog
PIK3	Phosphatidylinositide 3-kinases
rpm	revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
RNA	Ribonucleic Acid
SNP	Single nucleotide polymorphism
Tm	Melting temperature
TNM	Tumor-Node-Metastasis
UICC	International union against cancer
UV	Ultraviolet
µl	Microliters
µg	Micrograms
%	Percentage
>	Greater-than
≤	Less or equal to

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE LITERATURE REVIEW

1.1 Background

Breast cancer is one of the most common and potentially lethal diseases in women worldwide (Parkin, Pisani et al. 1999; Boyle, Leon et al. 2003). Differences in incidence and mortality in different populations and different regions of the world strongly indicate that the disease is multi factorial and both environmental and genetic factors are implicated (Cullen, Maguire et al. 2001; Nathanson, Wooster et al. 2001). Extensive research into the aetiology and pathobiology of breast cancer continues apace while other investigators are working on improving the detection methods. With reference to the latter, it is evident that the current detection techniques lack the desired sensitivity and specificity needed to ensure the early detection and ultimate survival of most cases of breast cancer and ultimately improve the survival (Tabar, Yen et al. 2003).

The breast consists of the mammary gland and associated skin and connective tissues. The mammary gland consists of a series of ducts and associated secretory lobules. These converge to form 15 to 20 lactiferous ducts, which open independently onto the nipple. The breast lies on deep fascia over pectoralis major muscle. With such anatomical location, it is related both to the thoracic wall as well as the structures associated with the upper limb; therefore the arterial supply, venous drainage as well as the lymphatic drainage occurs via multiple routes. The breast receives its blood supply from branches of the axillary, internal thoracic and intercostal arteries. Veins draining the breast parallel the arteries and ultimately drain into the axillary, internal

thoracic and intercostal veins. The lymph drainage of the breast is amazingly extensive; most of the lymphatic drainage occurs through the axillary nodes that receive approximately 75% of the lymphatic vessels. The supraclavicular, parasternal and abdominal lymph nodes as well as the opposite breast receive the remaining drainage (Richard 2009).

Breast disorders are very common and present mainly as benign or malignant breast lumps. Benign breast lumps are common in patients in the younger age group and can result from benign proliferations of either the glandular tissue, the ductal system or its fibrous surroundings. Histopathological and immunohistochemical examination of primary breast tumors has provided substantial evidence that breast cancer cells are epithelial in origin, originating either from the mammary lobules or ducts (Stevens, Lowe et al. 2002). Although evidence suggest that the growth and spread of these tumors usually depends on the exact cellular site of origin of the cancer, it is a known fact that metastatic breast cancer cells lodge mainly in the axillary lymph nodes. However, the extensive lymphatic drainage which may cause parallel spread of malignancy to other nodal groups makes the management of advanced breast cancer potentially difficult and eventually leads to treatment failure and death (O'Higgins, Linos et al. 1998).

Once a lump is detected, either by self examination, clinical examination, or screening, the next step is to establish whether such lump is malignant or not. This confirmation is accomplished by a biopsy followed by histopathological evaluation of the specimen. Generally, most breast lumps should be removed surgically and sent for pathological evaluation. If malignancy is confirmed, the next step is to attempt at finding whether metastatic cells have lodged in the regional lymph nodes . The search

for malignant cells will be then complemented by the search for metastases in organs most likely to harbour metastases like the lungs, liver and bones. This might be achieved via several detection means such as conducting hematological and radiological tests (Weis and Cheresh 2011).

Once this evaluation is complete the surgeon and the pathologist will be able to stage the tumor. In an attempt to achieve unity in describing solid tumors, the international union against cancer (UICC) has established a system that serves as a global language for the classification of solid tumors (Hermanek 1999). The system was called the TNM which stands for Tumor-Node-Metastasis. This system has been in existence for more than 50 years. It is based on measuring the size of the primary tumor, determining the number and the group of regional nodes involved and detecting the involvement of distant sites. Subsequently, these three variables will be assigned to four stages (I-IV) that describe the severity of the disease and help in defining the optimal treatment choices as well as providing a preliminary idea about the prognosis (Sobin, Hermanek et al. 1988).

The surgical management of breast cancer has evolved dramatically over the past century. In the beginning of the century, Halsted (1907) proposed that cancer progression followed a sequential pattern, and thus recommended the use of extensive radical surgery to cure it, as it was considered as a local-regional disease (Halsted 1907). The original procedure included removal of the breast and the underlying muscle together with an axillary nodal clearance. Such an invasive procedure was associated with many complications and slow recovery. Disappointed by the fact that radical mastectomy did not cure all the patients, other surgeons have reported their experience with less invasive surgical procedures and tried the preservation of the axillary nodes (Haffty, McKhann et al. 1993; Guenther, Hansen et

al. 2003). In an attempt to summarize the surgical approach in that period, Fisher (1999) has stated that *" Most of the period from 1900 to 1970 was governed by the 'non-science' of anecdotalism and clinical inductivism and was marked by the absence of science gestalt"*. Consequently, the shift in the paradigm of surgical intervention has appeared gradually, driven by the results of randomized clinical trials which have compared various techniques. For example, two of these trials were conducted by Fisher and co workers between 1971 to 1985. They reported that a 15 year follow-up of 2000 women who were randomly distributed among three treatment groups: total mastectomy, lumpectomy alone or lumpectomy followed by breast irradiation indicating that there was no difference in disease free- survival or overall survival rates. Meta-analyses of trials indicate equivalence of survival when various surgical techniques are compared. Thus, the current approach would be to perform a mastectomy for larger tumors, and wide local excision with radiotherapy for smaller ones, should these choices be acceptable to the patient (Fisher 1999).

The decision on whether to remove or preserve the axillary lymph nodes that are apparently free of clinically identifiable metastatic disease has changed over the years. When the tumor is smaller than two centimeters and the lymph nodes are not palpable, the initial surgical approach was to preserve the axillary nodes and complement the treatment with local radiation since the disease was initially thought to be locoregional (Dubois 1990). However, of concern was the report by Fisher et al (1983) in which it was stated that in 30% of women with node-negative axilla who underwent only mastectomy without lymphatic clearance, metastatic disease recurred within 5 years and eventually the woman died of it (Fisher, Bauer et al. 1983). Subsequently, other researchers have confirmed the importance of such observation. Of these was the report of Rosner et.al (1993) who have conducted a10-year follow-

up on 407 breast cancer patients that were treated by surgery alone reporting a 10-year recurrence rate of 19% (Rosner and Lane 1993). In addition, a recent report by Fodor et al (2005) after a 15 year follow-up, have stated that a relapse rate of 24% has been found in 92 breast cancer patients who had undergone mastectomy with nodal preservation and 75% relapse rate in 56 breast cancer patients who had only their primary tumor been removed. This clinical manifestation of distant relapse implies that these patients already developed sub clinical occult micro-metastases at the time of primary tumor excision and therefore the decision of preservation of axillary content is not necessarily a good decision (Fodor, Mozsa et al. 2005).

The above reports imply that as a precautionary measure, removal of the axillary lymph nodes should be carried out in each case. However, surgical experiences have revealed that complete nodal dissection is not without complications, since there is a high morbidity where many of the patients suffer from post mastectomy lymphedema. For example, Petrek et.al (2001) have shown that 49% of the 263 patients that had had complete nodal dissection developed lymphedema. For this reason, the search for methods to improve the detection of metastases while at the same time sparing the axillary contents became a necessity (Petrek, Senie et al. 2001).

Researchers reasoned that if metastasis had occurred, tumor cells might be detected in other areas of the body. Therefore, investigators begun to sample other areas such as blood and bone marrow (Domschke, Ge et al. 2013). Braun et.al (2000) have obtained bone marrow aspirates from both upper iliac crests of 552 patients with stage I, II, or III breast cancer and 191 patients with nonmalignant disease. The specimens were stained with the monoclonal antibody A45-B/B3, which binds to an

antigen on cytokeratins. The group has reported a detection rate of 36% (Braun, Pantel et al. 2000). Furthermore, Ballestrero et.al (2001) have studied the specificity of maspin mRNA marker when used in a nested polymerase chain reaction (PCR) assay to detect breast cancer micro-metastasis in the bone marrow. The marker was expressed in all the 41 breast cancer patients but also in 40% of the 17 patients with hematological malignancies other than breast cancer stating that the marker is not disease specific (Ballestrero, Coviello et al. 2001). Lin, Zhang et.al (2000) have tried sampling the blood and using a combination of 4 monoclonal antibodies (BM7, BM8 against MUC1, 5D3 against CK8,18,19 and HEA125 against human epithelial antigen) and a sensitive immunocytochemical staining procedure to identify breast cancer cells in leukapheresis products. However, their assay detection rate has not exceeded 20%. It has become evident after such studies that malignant cells are present in low concentrations in such samples which have made the attention to be redirected again towards examining the nodal basin (Lin, Zhong et al. 2000).

Sentinel lymph node biopsy (SLNB) has been introduced as an alternative method for evaluation of the lymphatic basin with further aim of proceeding with complete axillary lymph node dissection if the sentinel node has been found to contain metastases (Kiricuta 2000). The sentinel node was originally defined as the first node to receive lymphatic flow from a primary tumor. Subsequently, this definition has been revised as it has been found that in many instances, parallel lymphatic channels from the primary tumor drain to more than one node making such definition applicable to more than one node. To overcome such debate, the general surgical trend if more than one sentinel node can be detected at the time of surgical operation, is to remove all these nodes. In such cases the true sentinel node will be the node that contains metastatic cells. However, other nodes in same basins will not contain these

metastatic cells (Nathanson 1999). The development of sentinel node surgery has helped the pathologists to provide a more focused analysis. However, the extensive pathological work that has been dedicated towards improving the detection of metastases lacked standardization. This fact became evident when studies attempting to compare such results were confounded by the lack of consistency and subsequently, the prognostic impact of such micro-metastases remained questionable (Cserni 2004).

However, a very recent randomized clinical trial has shown that patients with limited sentinel lymph node metastastatic breast cancer treated with breast conservation and systemic therapy, the use of SLNB did not result in inferior survival (Giuliano, Hunt et al. 2011). In addition, a 10-year follow up of treatment outcomes in patients with early breast cancer and negative axillary lymph nodes that were treated with tangential breast irradiation following sentinel lymph node dissection or axillary lymph node dissection concluded that the conservative approach yielded a similar prognostic results (Wernicke, Goodman et al. 2011).

1.2 The history of BC cancer detection in South Africa

The following sections provide an insight with regard to the history of BC cancer detection in South Africa. A revision of the development of various molecular methods for optimal detection of genomic instability in patients with breast cancer has been included in this chapter.

1.2.1 Breast Cancer in South African Women: An updated report with an insight into Triple negative breast Cancer

In South Africa, this cancer follows similar worldwide statistics (Parkin, Sitas et al. 2008). National Cancer Registry (2013) shows that Breast Cancer is top of the list in South African females. A report on the cancer incidence in a rural population of South Africa has shown that breast cancer is the third common cancer after cancer of the cervix and oesophageal cancer. It was also documented that African women demonstrated a lower age of occurrence (35-40 years) when compared to other women globally (Somdyala, Bradshaw et al. 2010).

There are several known risk factors for BC in the African continent. BC risk is mainly attributed to menstrual and reproductive factors like the age at menarche, parity, and breast feeding and these vary according to the social class/status (Henderson, Ross et al. 1982) . In addition, obesity plays a major role as African women are known to have high body mass index (BMI) (Goedecke, Evans et al. 2011). Life style factors like high alcohol consumption, low physical exercise levels, and exposure to exogenous hormones either as contraceptives or postmenopausal hormone replacement therapy. However, these finding warrant further research in Africa (McCurtis Witherspoon and Williams Richardson 2006).

Among other risk factors, race seems to be an important determinant for the lifetime risks of developing breast cancer. Vorobiof et.al (2001) has shown that black women are less affected than white women and women of mixed ancestry.. Moreover, a recent report has indicated that breast cancer incidence ranges from a low occurrence of one in 81 in African women (similar to Japan) to as high as one in 13 among white women, similar to rates in Western countries (Vorobiof, Sitas et al.

2001). However, the myth of being a "disease of the white" has changed over the years (Helman 1977; Bradshaw and Harington 1982). It was documented that African black women with breast cancer have started to show an increasing incidence and tend to have a low survival rate. In 1984, Walker et al. reported on the poor survival of females affected by breast cancer but failed to indicate a possible aetiology (Walker, Walker et al. 1984; Walker, Adam et al. 2004). In addition, a report by Fregene et.al (2005) has indicated that the mortality rates are strikingly high among African women despite their low breast cancer incidence rate when compared to Western populations (Fregene and Newman 2005). Other reports have stressed on the fact that the disease is usually detected at its late stages (Stage III and IV) (Hassan, Onukak et al. 1992; Muguti 1993; Amir, Kitinya et al. 1994; Amir, Azizi et al. 1997). In addition the tumor grade and tendency for proliferation was reported to be higher in African women (Fregene and Newman 2005).

The high mortality in the Black African women might be attributed to several ongoing problems that are particular to the African continent. One of the most apparent problems is the lack of awareness of breast cancer which tend to be low on the list of health care priorities. A 2002 report has documented that 25% of urban South African women and 33% of rural women would seek the advice of a traditional healer when it comes to breast lumps (Pillay 2002). A second concern might be the access to breast cancer screening; as it was documented that rural black women have less access to screening than those living an urban life. This lack of medical services provision in South African women might be attributed to the socioeconomic status and other competing life threats like HIV/AIDS and tuberculosis (Brown and Williams 1994). All this have led to the fact that breast cancer is not detected at its early stages in the black community. It is beyond doubt that the early detection might

help in reducing mortality rate and improve survival in black women (Chu, Tarone et al. 1996).

Moreover, the validity of screening methods for the early detection of breast cancer has been questioned, despite the fact that early detection of breast cancer improves the therapeutic outcome (Meng, Ward et al. 2011). Recent studies have questioned the efficacy of conventional detection methods such as mammography and self-testing; self-testing has proven to have little to no effect on mortality rates (Thomas, Carter et al. 2002). On the other hand, mammography results in lower mortality rates in the population at large, but in young women the false positive rate results in a large number of unnecessary biopsies (Berry, Cronin et al. 2005). Therefore, the need has arisen to find other means for early detection of breast cancer.

The molecular screening of patients offers an accurate complimentary alternative to mammography. Unfortunately, despite the advance in molecular methodologies, the genetic profiles associated with breast cancer in Africa remains poorly investigated. This is attributed mainly to the cost of such studies and the lack of proper documentation of patients clinical data (Basro and Apffelstaedt 2010). However, in the last two decades, several studies have tried to unravel the molecular event associated with breast cancer (Sjogren, Inganas et al. 1996; Hilakivi-Clarke, Wang et al. 2004; Fisher, Cole et al. 2010).

In addition, Ameyaw, Thornton et.al (2000) discovered that the over-expression of protooncogene Her2/ neu which is reported in 15-30% of breast cancers was also studied in African women; blood from 200 Ghanaian , 257 Caucasian and 90 African American healthy women was genotyped. The known single nucleotide polymorphism (Val655Ile) which is associated with an aggressive phenotype was not

found in the African specimens, but it was found in 20% of the Caucasians and 24% of the African-Americans recruited for the study (Ameyaw, Thornton et al. 2000).

However, in the last few years, two interesting articles have emerged. The first study done by Reeves et.al (2004) has reported on the role of BRCA1 in South African breast and/or ovarian cancer families. A total of 90 South African breast/ovarian cancer families were screened for BRCA1 mutations by means of PCR-based mutation detection assays. Eighteen families (20%) were identified with BRCA1 disease-causing mutations. Four Ashkenazi Jewish families were identified with the 185delAG mutation, whereas 2 Afrikaner and 1 Ashkenazi Jewish family were found to harbor the 5382insC mutation. Five of the families (5.56%), all of whom are Afrikaners, were found to carry the novel E881X mutation. Genotype analyses show that these patients share a common ancestor. Genealogic studies have identified 3 possible founding couples for this mutation, all of whom arrived in the Cape from France in the late 1600s. Of the remaining mutations detected, 3 have not been reported previously and include the S451X, 1493delC and 4957insC mutations (Reeves, Yawitch et al. 2004). A consecutive study has recruited 105 Coloured and 16 Black Xhosa women residing in the Western Cape of South Africa diagnosed with breast cancer. These patients were screened using a panel of six previously reported SA Afrikaner and Ashkenazi Jewish BRCA1/2 mutations and identified only two Afrikaner mutations. In addition, screening by the protein truncation test (BRCA1 exon 11, and BRCA2 exons 10 and 11) revealed an additional four deleterious mutations(BRCA1c.1504_1508del,p.Leu502AlafsX2,BRCA2c.2826_2829del,p.Ile943LysfsX16,c.6447_6448dup,p.Lys2150IlefsX19 and c.5771_5774del,p.Ile1924ArgfsX38). The latter, also known in Breast Cancer Information Core nomenclature as 5999del4, was identified in 4 of 105 (3.8%) Coloureds and 4 of 16 (25%) Xhosa

women, which makes it a frequent mutation in the Western Cape Province. Despite the fact that the previous study has reported that this mutation occurs in the Netherlands, haplotype analysis indicated two distinct origins for the Dutch and South African mutations, excluding the possibility of a common Dutch ancestor and suggesting gene flow from the indigenous tribes such as the Xhosa to the Coloured population instead (van der Merwe, Hamel et al. 2012). Currently, these mutations are linked to the triple negative phenotype breast cancer. The reasons for these associations are unclear but may ultimately provide avenues for prevention and targeted therapy (Irvin and Carey 2008).

It is well-known that when the cancer is discovered early, patients live longer and require less extensive treatment (Lurie and Parker 2007). For the previously mentioned reasons, treatment options were limited, most of these women were exposed to mastectomy and its post-operative complications due to the advanced stage of the disease where breast conservation was not an option (McNeely, Binkley et al. 2012).

1.3 Triple Negative Breast Cancer (TNBC)

1.3.1 Is TNBC common in South Africa? Research attempts to sub-classify TNBC cases

Until recently, the hormone receptor status was not investigated in most parts of Sub-Saharan Africa simply because of the lack of technology, therefore, many women were treated blindly with Tamoxifen, the known hormonal treatment for breast cancer, and this approach was acceptable somehow because of the minimum side effects (Clegg-Lamptey, Dakubo et al. 2009). Subsequently, several groups of researchers have started to report on these receptors' absence and worked towards

finding better histological sub-types of these tumors with the further aim of finding means of alternative treatment approaches and gaining a better insight regarding the disease prognosis (El Khouli, Macura et al. 2009).

Within the last decade, the concept of triple negative breast cancer (TNBC) was introduced and refers to approximately 10–15% of breast carcinomas . TNBC is defined as a subtype that lacks the expression of oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). The treatment of these tumors is challenging because there is no specific target for hormonal treatment. However and despite the fact that TNBC exhibits poor response to hormonal treatment, many reports have indicated that such tumors display increased chemo sensitivity when compared with other breast tumor subtypes (Chavez, Garimella et al. 2010).

In order to refine the choice of treatment, there was a need to define the heterogeneity of TNBC. Several ongoing studies are trying to characterize these tumors at the molecular level. Studies based on the proliferation-related marker Ki 67 have shown that at least two subgroups exist within TNBC, each with a different response to chemotherapy and ultimately they varied in prognosis (Giuffre, Adamo et al. 2011).

Researchers have initiated immunohistochemical studies in an attempt to get a better means of classifying patients and finding means for treatment. Initially, studies based on basal immunohistochemical markers (epidermal growth factor receptor type 1 (EGFR) and cytokeratin 5/6 have shown that at least two subgroups exsist (Giuffre, Adamo et al. 2011). More recent studies like the report by van Bogaert (2013) have identified 6 molecular subtypes of TNBC using immunohistochemistry. These were

the Luminal A(ER+/PR+, HER2-), luminal B(ER+/PR+, HER2+), triple negative (ER-/PR-, HER2-), basal like(ER-/PR-, HER2-, CK5/6+), null type (ER-/PR-, HER2-, CK5/6-) and tumors that over express HER2 which are usually associated with poor prognosis. He concluded that there is a commonality between the various subtypes among African-American and women in sub-Saharan Africa (van Bogaert 2013). Another study has reported a similar classification depending on the gene expression profiling of 587 TNBC cases (Lehmann, Bauer et al. 2011).

It was reported that TNBC has a worse prognosis and tends to relapse early compared with other subtypes of breast cancer (Dawood 2010). Several prognostic markers were studied to get an insight about the prognosis. Of these the Ki-67 was studied and its high level is associated with favourable prognosis (Nishimura, Osako et al. 2010) . Furthermore, gene expression studies have classified TNBC into two main subtypes, claudin low and basal like with no major prognostic difference. The latter have been found to have higher percentage and is associated with P53 mutations. In addition, the overlap between TNBC and BRCA1 and BRCA2 mutations was also reported. These tumors are reported to have a favourable prognosis. This heterogeneity in classification made comparison between different reports difficult and warranted further studies (Perou 2010).

1.4 Pathological and Clinical characteristics of TNBC

Research within the pathology department worldwide has attempted to find clear pathological stratification on the basis of the difference in gene expression profile Five major groups were characterised as follows: the first two groups are called Luminal A and B and were ER positive. The remaining three were ER

negative and grouped as follows: A“Basal-like breast cancer characterised by the lack of expression of ER, PgR and HER-2 as well as incresead expression of basal cytokeratins CKs 5/6 and 17. The second type is erbB2 like/HER-2 like with high expression of erb2 and thirdly the normal like BC showing molecular charecteristics of normal tissue (Sorlie, Perou et al. 2001).

Most TNBC have demonstrated a very aggressive phenotype with almost 80% clinical and pathological overlap with BRCA1 and basal-like subtype of BC. It is documented that more than 90% of TNBC demonstrate an invasive ductal histology, with high mitotic index, central necrotic zones and pushing borders as well as lymphocytic infilterate (Gluz, Liedtke et al. 2009).

Clinically, most TNBC patients have an early onset of presentation. However, a team has reported on the detection of TNBC in older women which contradict the fact of favourable prognosis being associated with late presentations and calls for the need of more serious interventions in the older age group (Aapro and Wildiers 2012) .

The mortality associated with TNBC is alarming in women of African descent. Metastases happen within the first three years of treatment. In addition, the 5-year survival rate is much lower when compared with other types of breast cancer. Thus there is a continuous search for early detection-markers of prognostic significance that might help in improving the prognosis (Boyle 2012).

The fact that these tumors do not respond to the known hormonal treatment like Tamoxifen and Trustazumab had led to the search for novel therapeutic targets. Several agents are currently being investigated as potential therapeutic agents for the treatment of women with TNBC including agents targeted against epidermal growth factor receptor *EGFR*, poly (ADP-ribose) polymerase (PARP) inhibitors, anti-

angiogenic agents, and multityrosine kinase inhibitors showing promising results (Anders and Carey 2008; Rottenberg, Jaspers et al. 2008).

1.5 Methodologies to identify molecular markers of early changes.

1.5.1 Genomic instability as a means for early detection of breast cancer

It is known that breast cancer exhibits a heterogenous behaviour with regard to metastatic potential and response to treatment (Silva, Dominguez et al. 1999). Accumulating body of evidence has shown that the study of early genetic alterations in tumor cells seems to help in identifying markers of prognostic significance.

Advances in methodology have allowed us to more precisely determine the approximate chronology of some of these aberrations and the possible roles each plays in the formation of malignancy. It is now proposed that the early loss of cell cycle control in the presence of a mitogenic stimulus allows a cell to divide unchecked. Such uncontrolled proliferation would yield a high level of genomic instability. As proliferation continues, numerous additional chromosomal abnormalities occur, and increased tumor heterogeneity would be observed as distinct subpopulations emerge in the evolution toward a progressively more aggressive phenotype. However, much still remains to be learned to gain a full understanding of the key players behind the genetic evolution of breast cancer. This can be only achieved by analyzing the pre-invasive and putative early stages of breast cancer in order to characterize the most probable sequence of genomic abnormalities (Brenner and Aldaz 1997).

Two methods for the detection of changes in the DNA of tumor cells have been explored in this study.

1.6 Loss of heterzygosity/allelic instability in patients with BC:

The allelic losses at particular chromosome regions are common (Isaacs, Cavalli et al. 2004). These losses might indicate deletion in the second allele of a tumor suppressor gene. However, this phenomenon might be due to other mechanisms associated with these deletion like aberrant somatic recombinations and mitotic nondisjunction (Murthy, DiFrancesco et al. 2002). The allelic instability has been investigated in patients with TNBC. A study has analysed 22 common breast cancer susceptibility variants in 2,980 Caucasian women with triple-negative breast cancer and 4,978 healthy controls. The team had identified six single-nucleotide polymorphisms, including rs2046210 (ESR1), rs12662670 (ESR1), rs3803662 (TOX3), rs999737 (RAD51L1), rs8170 (19p13.1), and rs8100241 (19p13.1), these were significantly associated with the risk of triple-negative breast cancer (Stevens, Fredericksen et al. 2012).

The detection of LOH in DNA material obtained from ductal lavage (DL) and nipple aspirate fluid (NAF) in BC patients have generated promising data. A goup have studied 26 DL and 6 NAF samples from 14 women of known BRCA1 status who have no evidence of breast cancer: 9 mutation carriers and 5 non-carriers. LOH studies for BRCA1 were possible in 19/26 DL samples (Isaacs, Cavalli et al. 2004). This has provided evidence for the feasibility of performing molecular studies on DNA obtained from DL and NAF (Isaacs, Cavalli et al. 2004).

Individual studies of LOH has rarely spanned more than one chromosome. However, a pooled analysis of LOH through a complete genome scan has identified several tumor suppressor genes through the entire genome: A group of researchers has compiled an extensive database from 151 published LOH studies of BC with summary data of more than 15000 tumors and had shown a striking preferential loss in specific regions of chromosomes 7q, 16q, 13q, 17p, 8p, 21q,3p,18q,2q and 19p as well as other regions (Miller, Wang et al. 2003).

Allelic loss at BRCA1, BRCA2, and *p53* seems to be a frequent event in sporadic BC particularly at younger age. Several groups have reported on the association between sporadic cancer and loss of heterozygozity in BRCA1, BRCA2, and *p53*in sporadic cases of breast cancer (Johnson and Lo 2002; Charef-Hamza, Trimeche et al. 2005). The interferon regulatory factor-1(*IRF-1*) gene localized on chromosome 5q31.1. is believed to act as a tumor suppressor gene in breast cancer through activities associated with caspase activation and apoptosis induction (Bouker, Skaar et al. 2005). Kim, Armstrong et al (2004) has reported that the expression of *IRF-1* induces apoptosis and inhibits tumor growth in mouse mammary cells (Kim, Armstrong et al. 2004). The *IRF-1* was found to be inactivated by a single nucleotide polymorphism (SNP) in BC. This might indicate that the loss of its function is critical to the development of the disease. Furthermore, the low *IRF1* mRNA expression has been strongly correlated with high risk of recurrence and death (Cavalli, Riggins et al. 2010).

1.7 DNA copy number changes in the genes of The phosphatidyl inositol 3-kinase/AKT signalling pathway and its role in early breast cancer:

Accumulating evidence indicates that the phosphatidyl inositol 3-kinase (*PIK3)/ATK* pathway (as illustrated in Figure 1.) facilitates a number of extracellular signals and impacts on several cellular events like cell growth, proliferation and survival (Dillon, White et al. 2007). It is believed that the (*PIK3)/ATK* is relatively an early event in breast tumorigenesis preceeding invasion. The controlled signalling in cancerous cells usually happen through mutation, amplification of the Tyrosine Kinase or *PIK3* itself (St-Germain, Gagnon et al. 2004). The *(PIK3)/AKT* signalling is initiated by the growth factor binding to the tyrosine kinase receptor leading to receptor dimerization. As a result, lipid kinase *PIK3* is recruited to the internal docking site and becomes activated. *PIK3* next converts membrane lipid *PIP2* to its active *PIP3* form which in turn leads to the activation of key signalling kinase, *AKT*. *AKT* promotes cell growth through protein synthesis driven by mTOR signalling and reduces cell death by blocking FOXO activity (Carracedo, Ma et al. 2008; Zhang, Fang et al. 2010; Logue and Morrison 2012) .

Aberrant activation of this pathway has been implicated in many cancers as well as resistance to cancer therapies. In breast cancer, aberrant expression of this pathway is reported in 50% of the cases and mutation in the alpha catalytic subunit of phosoinositol-3-kinase *(PIK3CA)* occurs in approximately 30% of ER Positive breast cancers (Ellis, Lin et al. 2010). A group of researchers has examined the later mutation response to neoadjuvant endocrine therapy for ER positive breast cancer (Ellis, Miller et al. 2009) and found a weak negative correlation between the *PIK3CA* mutation status in those patients and their clinical response to treatment. The study also documented a favourable relapse free survival rate in those patients but failed to

explain the mechanisms behind the improved prognostic indicators (Ellis, Miller et al. 2009).

The frequent dysregulation of the *PIK3/AKT* pathway in breast cancer has made the signalling players of this pathway attractive for therapeutic targeting. Nevertheless, a further insight and a more comprehensive evaluation is needed to unravel the mysteries behind the signalling intricacies (Dillon, White et al. 2007). This will help to develop therapeutic agents to target not only molecules but functional aspects of breast cancer progression (Suter and Marcum 2007).

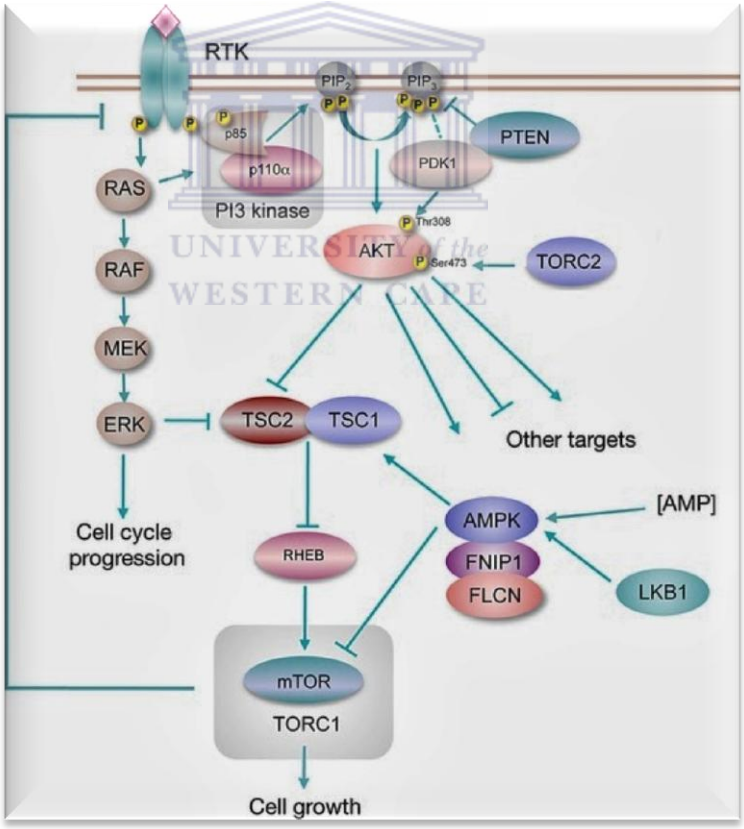


Figure 1.1: The main genes involved in the *PIK3* pathway (Holmes 2011)

Aims of the study

The aims of this study were:

- To detect and report on the loss of heterozygosity / allelic instability of *IRF-1* in breast cancer cases from South African women.
- To correlate LOH of the *IRF-1* gene with clinical characteristics of the patients (age, stage and ER status).
- To detect and report on the genetic alteration in the *PIK3* pathway in breast cancer cases from South African women.
- To correlate the alterations observed in the TNBC versus non-TNBC cases.

Therefore, this study was designed to identify genes associated with the triple negative tumors that serve as biomarkers for early detection, prognosis as well as targets for treatment.

CHAPTER 2

METHODOLOGY

2.1 Tumor and Normal Tissue Specimens

The specimens were supplied by the National Health Laboratory Services in Cape Town at the Division of Anatomical Pathology (UCT). Subsequently, ethical approval was obtained UCT and UWC (refer to Appendix VIII) . A total of 101 patients were randomly selected for this study. A formalin fixed paraffin embedded block with breast tumor tissues and a second block with normal tissue were collected for each case from the archives of the National Health Laboratory Services in Cape Town at the Division of Anatomical Pathology (UCT). The clinical data pertaining to age, race, tumor stage, and receptor status were obtained for each patient from the department of surgical pathology at Groote-Schuur Hospital. Patients' identities were kept anonymous at all stages of the project

2.2 Patient characteristics

The median age of the study population at diagnosis was 53 years (range, 22-80 years). Seventy nine (79) patients were in an early tumor stage and 22 patients in late tumor stage. The details on the distribution of clinicopathological factors in the study cohort are given in Table 2.1. Refer to Appendix VI for the summary of patients's clinical data.

Table 2.1: Clinicopathologic characteristics of 101 breast cancer patients

Characteristics	Patients, no. (%)
Age, yrs	
≤50	43 (42.57)
>50	58 (57.420
Tumor stage	
Early (stage I&II)	79 (78.21)
Late (stage III&IV)	22 (21.78)
Hormone receptor status	
ER negative	45 (44.55)
ER positive	56 (55.44)
PR negative	98 (97.0)
PR positive	3 (2.97)
HER2 negative	57 (56.44)
HER2 positive	29 (28.71)
HER2 equivoeal	13 (12.87)
TNBC	27 (26.73)

2.3 Histopathological examination and Micro-dissection of tissue

In most cases, 3 consecutive 10-µm sections from each block were prepared, of these one was Haematoxylin & Eosin (H&E) stained. Stained slides were initially examined by a pathologist for the presence of breast cancer cells. Once found, the positions of these cells were marked and the consecutive slides were then micro-dissected with a surgical needle to isolate the malignant cells to ensure pure tumor cell population. Dissected tissue was then placed in 1.5ml micro centrifuge tubes. For each case, a similar procedure was followed to obtain a normal tissue for the same patient from a different block.

2.4 DNA Isolation from paraffin-embedded tissue

2.4.1 Removal of paraffin from the tissue

For each case, the tissue was washed three times with 100µl of Xylene for five minutes to dissolve the wax. To concentrate DNA samples, the ethanol precipitation method was used, therefore the tissue was washed three times with 100µl of 100% Ethanol for five minutes and was left to dry. A volume of 180 µl of ATL buffer was added. (DNAeasy, Qiagen Inc. Valencia,Carlifonia), followed by 20µl Proteinase K. The micro centrifuge tubes were put in a 55°C shaking heat block and left overnight until all the material was dissolved. Proteinase K was inactivated by heating the samples to >95°C for 10 minutes using the Thermal Cycler. Samples were left to cool at room temperature. To cleave any RNA present and ensure a pure extraction of DNA, free from RNA impurities 0.005 RNase A (0.6µ was used) stock solution (20µg/µl) was added to achieve a final concentration of [RNase A] =0.1µg/µl. The samples were incubated at 37°C using the heat block for 1.5 hours and were left to cool at room temperature. 1 volume of Phenol: chloroform: isoamyl: alcohol (25: 24: 1) was used to remove proteins from the preparation of nucleic acids and vortexed for 15 seconds and centrifuged at 14 000 rpm for 5 minutes. The aqueous (top) layer was transferred to a new tube. One (1) volume of Phenol : chloroform : isoamyl alcohol (25: 24:1) was added again, vortexed for 15 seconds and centrifuged at 14 000 rpm for 5 minutes. The aqueous (top) layer was transfered to a new tube. A 0.2 volume of 10M Ammonium acetate (NH_4OAc) to achieve final [NH_4OAc]=2M was added. A 0.08 volume of glycogen (8µl used) was added and 2 x volumes of 100% Ethanol was added (200µl used). The samples were left at -20°C freezer overnight. The next day the samples were centrifuged at 14 000 rpm for 20 minutes. The supernatant was

removed while being careful to retain the DNA pellet. The pellet was rinsed with 1ml 70% ethanol and the samples were centrifuged at 14 000 rpm for 5 minutes. Ethanol was removed and the pellet was left to dry. The pellet was re-suspended in 50µl of T.E (30µl of DNA Hydration solution was used) and the tubes were incubated at 65°C using the shaking heating block for two hours.

2.5 Quantification of DNA

The purity and yield of the total DNA was measured spectrophotometrically at 260 and 280 nm using a NanoDrop 1000 (Thermo scientific, USA). The presence of DNA was checked by gel electrophoresis (Appendix Ia and Ib).

2.6 Polymerase chain reaction (PCR)

Aliquots of DNA (2µg) were transferred to 25µl reactions containing final concentrations of 1X PCR buffer, 1.5 mM $MgCl_2$, 10 µM of each upstream and downstream primer, 0.1 mM of each dNTP, 1.5 U of GoTaq DNA polymerase and nuclease free water. All reagents were from Whitehead Scientific (Cape Town, South Africa) except when otherwise indicated.

PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK) as follows: The first PCR cycle consisted of denaturation at 94°C for 5 minutes. This was followed by 35 cycles at 94°C for 30 seconds, 55°C for 30 seconds (the annealing temperature for the *IRF-1* gene) and 72°C for 30 seconds, then a final extension at 72º C for 10 minutes. Optimization of the annealing temperature was carried for the IRF1 gene to select the optimal annealing temperature for this gene.

PCR products were separated by electrophoresis on a 2% agarose mini-gel (containing 0.5 µg/ml ethidium bromide). Thereafter, the bands of interest were analyzed with the Doc-ItLS® Image Acquisition Software, Version 5.5.5a (Doc-It Life Sciences Software,UVP™, Inc., San Gabriel, USA) and viewed under a UV Transluminator (UVP™, Inc.,San Gabriel, CA 91778, USA). A positive result was recorded when a band of the appropriate size (187bp) as ascertained by comparison with a 100-bp ladder was visible on the gel. (refer to Appendix III). Opening of the tubes containing PCR products and gel electrophoresis were always performed under negative atmospheric pressure to prevent subsequent amplicon contamination of DNA samples or RT-PCR reagents. The sequences of the oligonucleotide primers were obtained from the Genome Data Base (GDB) (http://www.gdb.org): Forward: 5′-ATGGCAGATAGGTCCACCGG-3′/ Reverse: 5′-TCATCCTCATCTGTTGTACG-3′ (Appendix II). Primers were fluorescently labeled. Synthesis and purification of each primer set was done by Whitehead Scientific (South Africa).

Table 2.2: Primer pairs amplicons lengths, sequences and melting temperature of the IRF1 gene: Evaluated by conventional RT-PCR. Primer position on the sequence (Appendix II).

Gene	Product size	Gene Bank Accession Number	Primers Sequence	Tm
IRF1	**187bp**	NC_000005	Forward: 5′-ATGGCAGATAGGTCCACCGG-3′/ Reverse: 5′-TCATCCTCATCTGTTGTACG-3′.	**F 55 ºC R 55 ºC**

2.7 Loss of Heterozygosity (LOH) analysis

To study LOH at the *IRF-1* locus, we selected an intragenic, dinucleotide, polymorphic marker (*IRF1* Dinucleotide Repeat, Allele Set GDB: 211036), with a high degree of heterozygosity (74% heterozygosity). The sequences of the oligonucleotide primers were obtained from the Genome Data Base (GDB) (http://www.gdb.org): Forward: 5′-ATGGCAGATAGGTCCACCGG-3′/Reverse: 5′-TCATCCTCATCTGTTGTACG-3′. Primers were fluorescently labeled and PCR amplification was performed using a standard protocol. Allele sizes were determined by electrophoresis of PCR products in 6% denaturing polyacrylamide gels and were compared to ROX 500 size standards (Applied Biosystems, Foster City, CA), using an automated sequencer (ABI 377), according to the manufacturer's instructions (Applied Biosystems). Fluorescent signals from the different size alleles were recorded and analyzed using GENOTYPER version 2.1 (Appendix IV) and GENESCAN version 3.1 software (Applied Biosystems). Following visual examination of computer printouts by two independent observers, LOH was determined mathematically according to the Genotyper User's Manual (Applied Biosystems).

2.8 DNA copy number assay:

TaqMan copy number assays (Applied Biosystems, CA), to identify gene copy number changes in the PIK3 genes *(ATK1, ATK2, PIK3CA, PTEN, FGFR2)* were performed in the microdissected DNA isolated from the FFPE tumor sections and from the normal control (pool of DNA from multiple female donors with no cancer). The samples were diluted to 5ng/µl and amplified in the ABI 7500 HT Real Time

PCR instrument (Applied Biosystems, Foster City, CA, USA) using *PIK3* specific probes and primers. The *RNASE P* gene was used as a reference. The gene sequences were based on the hg19.v9 genomic database. The probe locations of the *PIK3* and *RNASE P* genes were as follows: ATK1 105241518, ATK2 40761052, PIK3CA 178876317, PTEN 89685514, FGFR2 123325143 and 157907524 (14q11.2), respectively. *PIK3* and *HER2/NEU* probes were labeled with the fluorescent dye FAM and *RNASE P* with VIC dye (detailed information of the probes used and PCR conditions are shown in Table 2.3). PCR was performed in a total volume of 10µl in each well of a 96-well plate, which contained TaqMan Universal MasterMix (Applied Biosystems), 5ng of genomic DNA and 12.5 picomoles per liter concentration of each primer and probe. PCR conditions included an initial denaturation step of 95°C for 10 minutes, followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. All of the reactions were carried out in triplicates, and a negative control with no DNA template was included in every PCR run. For all PCR assays, crossing threshold (C_t) numbers were established using SDS 2.2 RQ software (Applied Biosystems), and the copy numbers were normalized against the reference gene *(RNASE P*) and determined by the $2^{-\Delta\Delta Ct}$ method. Copy number changes were only considered for samples that presented CT cycles $\leq$ 33 and a Z-score value $\leq$ 2.65. A two-fold increase or decrease in the copy number of *PIK3* genes in tumor samples in comparison to the normal control was considered as amplification or deletion, respectively.

Table 2.3: Probes and PCR cycle used for the TaqMan Copy number assay.

		TaqMan Probes	
		Position	Labeling
Target Genes	*ATK1*	105241518	5'=Reporter-FAM 3'=Quentcher- **nao** fluorescent
	ATK2	40761052	
	PIK3CA	178876317	
	PTEN	89685514	
	FGFR2	123325143	
Reference gene	RNASE P	157907524	5'=Reporter-VIC 3'=Quencher-TAMRA
Master Mix	Taq Gold DNA polymerase (Hot start), dNTPs		
PCR cycle	Hold: 95ºC-10 minutes		
	40 cycles:		
	95ºC- 15 seconds/60ºC- 60 seconds		

CHAPTER 3

RESULTS

3.1 Loss of heterozygosity (LOH) analysis for *IRF-1*

LOH is the loss of one allele at a specific locus, caused by a deletion, mutation; or loss of a chromosome from a chromosome pair, resulting in abnormal hemizygosity. This is detected when heterozygous markers for a locus appear monomorphic due to the deletion of one of the alleles. When this occurs at a tumor suppressor gene locus where one of the alleles is already abnormal, it can result in neoplastic transformation (Fujii, Ajioka et al. 2002).

In this study, 46 cases (45.5%) out of 101 cases were informative for the *IRF-1* dinucleotide marker used for LOH analysis. Figure 3.1. shows that LOH was detected in 23 of these informative cases (23/46; 50%). There was no statistical significance between LOH at the *IRF-1* locus and both older age (P value = 1.0000) and earlier stage (Stages I and II) (P value= 0.4982) based on Fisher's exact test.

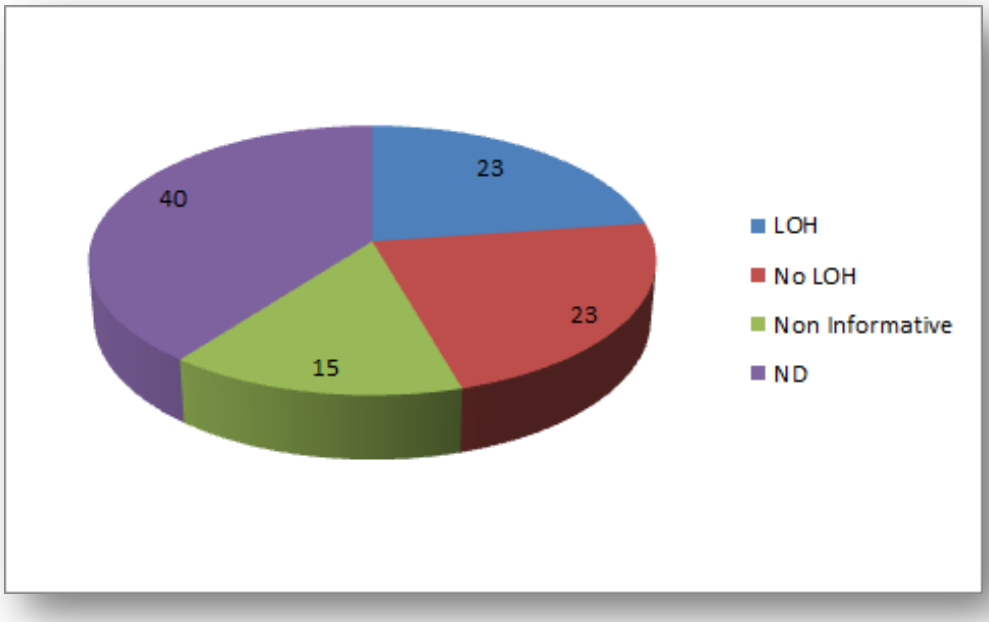


Figure 3.1: Pie chart representing cases of LOH at the *IRF-1* gene on 101 samples. 23 cases with loss of heterozygosity (LOH), 23 cases with no loss of heterozygosity (No LOH), 15 cases that were non-informative (NI) and 40 cases that were non dictatable (ND).

Figure 3.2 dipicts the PCR results for the amplification of the IRF-1 gene of cases 1 to 11 on tumor tissue samples and the remainder of the samples are shown in Appendix III.

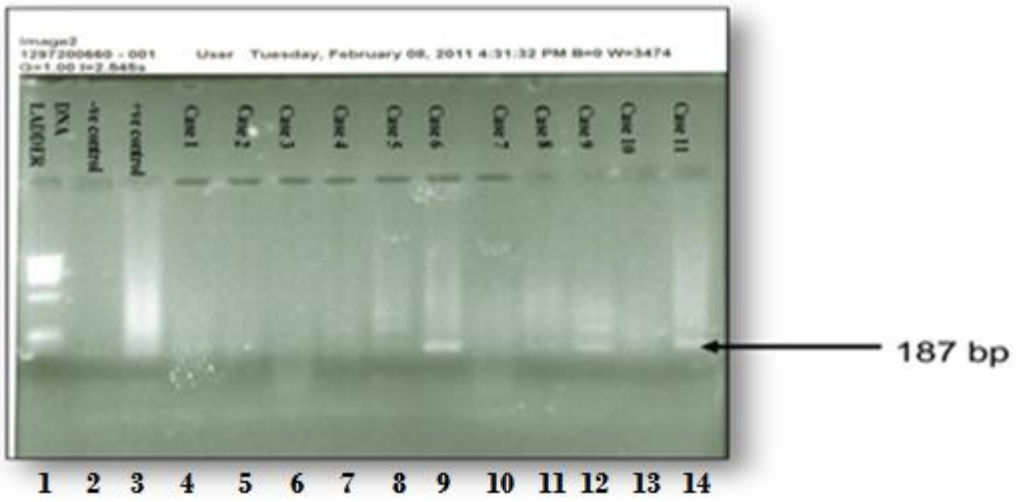


Figure 3.2: PCR amplification of the *IRF-1* gene of Cases 1-11 on tumor tissue. Lane 1: 100 bp DNA Marker; Lane 2: Negative control; Lane 3: Positive control; Lane 4: sample 1; Lane 5: sample 2; Lane 6: sample 3; Lane 7: sample 4; Lane 8: sample 5; Lane 9: sample 6; Lane 10: sample 7; Lane 11: sample 8; Lane 12: sample 9; Lane 13: sample 10; Lane 14: sample11. The arrow indicates expression of IRF1 at 187 bp.

Figure 3.3 depicts the detection of loss or no loss of heterozygosity, and the remainder of the samples are shown in appendix IV a, b & c. The y axis represents allele frequency and the x axis represent the allele base-pair length. In Figure 3.3, **A** represents a case with LOH and **B** represents a case with no LOH. On the figure ***a*** and ***b*** *Top panels* represent the analysis performed in normal cells and the *bottom panels* represent the analysis performed in tumor cells.

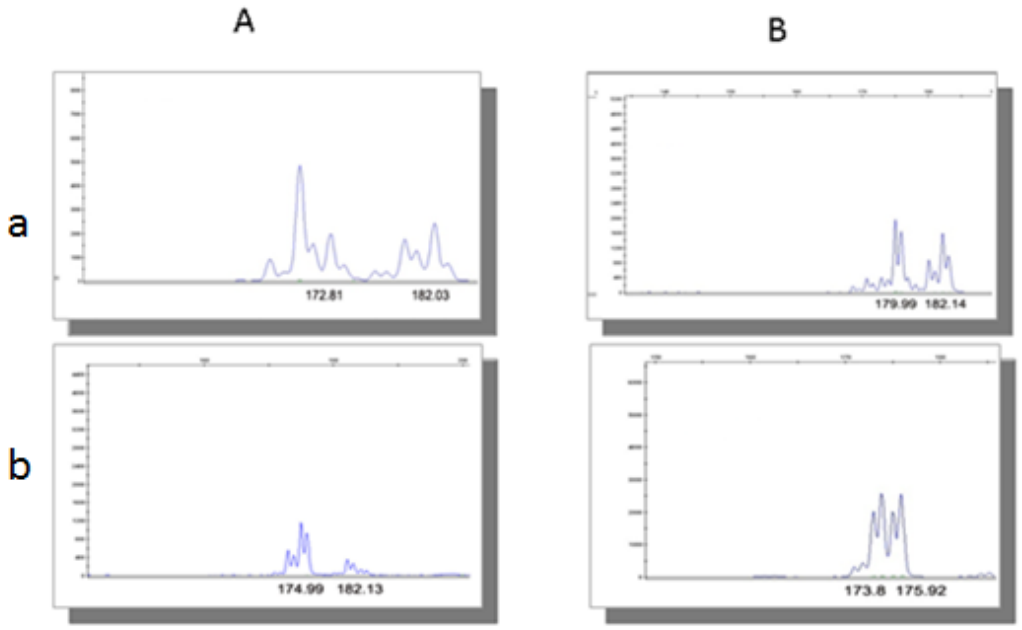


Figure 3.3: LOH analysis at the *IRF-1* locus of two representative breast cancer cases. **A** on the figure represents a case with LOH. **B** represents a case with no LOH. In **a**, **b**, *Top panels* shows the analysis performed in normal cells and *bottom panels* the analysis performed in tumor cells. The y axis represent allele frequency and the x axis represent the allele base-pair length

3.2 Correlation between loss and no loss of heterozygozity with other parameters

3.2.1 Correlation between loss and no loss of heterozygozity with age.

The striking correlation between advanced age and an increased incidence of cancer has led investigators to examine the influence of aging on genome maintenance. Since the loss of heterozygosity (LOH) can lead to the inactivation of tumor suppressor genes, and thus carcinogenesis, understanding the effect of aging on this type of mutation event is particularly important. There are several factors that may affect the rate of LOH, these include (i) an increase in the amount of DNA damage; (ii) specifically double-strand breaks (DSBs), and (iii) the ability to

efficiently repair this damage via pathways that minimize the loss of genetic information. Because of experimental constraints, there is only suggestive evidence for a change in the rate of DNA damage as humans age (August, Rea et al. 1994).

However, recent studies in model organisms find that there are increased rates of LOH with age, and that repair of DNA damage occurs via a different pathway in old cells versus young cells. It is speculated that the age-dependent change in DNA repair may explain why there is increased LOH, and that the findings from these model organisms may extend to humans (Carr and Gottschling 2008).

Table 3.1 depicts that there is no statistical significance (P=1.0000) between age (≤50 years or >50 years), LOH and no LOH. The results in the table show cases with LOH and are ≤50 years (23.9%), cases with no LOH and ≤50 years (26.08%), cases with LOH and >50years (26.08%) and cases with no LOH >50 years (23.91%).

Table 3.1 The correlation of LOH, NO LOH with ages ≤ 50 and > 50 years.

	Age		
	≤50 years	>50 years	**TOTAL**
LOH	11 (23.91%)	12 (26.08%)	23
NO LOH	12 (26.08%)	11 (23.91%)	23
TOTAL	23 (49.99%)	23 (49.99%)	46

P value =1.0000

3.2.2 Correlation between loss and no loss of heterozygozity with tumor stage

The development of human cancer is generally thought to entail a series of events that cause a progressively more malignant phenotype. This hypothesis predicts that tumor cells of the ultimate stage will carry each of the events, cells of the penultimate stage will carry each of the events less the last one and so on. That is to say, a dissection of the pathway form a normal cell to a fully malignant tumor may be viewed as the unraveling of a nested set of aberrations (Cavenee 1989).

Most LOH studies of breast cancer provide limited or no specification of tumor stage or histology, presenting an additional challenge in comparing across studies. For all of these reasons, previous summaries of LOH in breast cancer have been largely qualitative, consisting of descriptions of chromosomal arms or regions reported frequently lost (Lasko, Cavenee et al. 1991). Wang et.al (2003) evaluated the potential utility of occult circulating tumor DNA as a molecular marker of disease

in subjects previously diagnosed with breast cancer, and found no association, between plasma LOH and tumor stage or clinical status at time of blood collection (Wang, Larson et al. 2003).

In this study the results in Table 3.2 indicate that there is no statistical significance (P value = 0.4982) between loss and no loss of heterozygozity with tumor stage. This was observed from 44 samples which indicated that 16 (15.8%) cases had LOH,16 (15.8%) had no LOH and both groups were at early stages (I and II). Eight samples (7.9%) had LOH, four (4%) had no LOH and both groups were at a late breast cancer stage (III and IV).

Table 3.2: The correlation of LOH, NO LOH with early and late stages.

	Stage		
	Early stage	**Late stage**	**TOTAL**
LOH	16 (15.8%)	8 (7.9%)	24
NO LOH	16 (15.8%)	4 (4%)	20
TOTAL	32 (31.6%)	12 (12%)	44

P value = 0.4982

3.2.3 Correlation between loss of and no loss of heterozygozity with Estrogen receptor gene

Table 3.3 indicates that there is a statistical significance (P=0.0009) between loss and no loss of heterozygozity with ER gene. This was observed from 37 samples which indicated that 15 (14.9%) cases had LOH, 11 (10.9%) had no LOH and both groups were ER+, No (0) samples had LOH and 11 (10.9%) had no LOH and were ER-.

Table 3.3: The correlation of LOH, NO LOH with Estrogen receptor gene (+ve/-ve).

	Estrogen receptor		
	ER +ve	**ER -ve**	**TOTAL**
LOH	15 (14.9%)	0 (0%)	15
NO LOH	11 (10.9%)	11 (10.9%)	22
TOTAL	16 (26%)	11 (11%)	37
		P=0.0009	

3.4 DNA copy number changes on specific genes in the PIK3 pathway

3.4.1 DNA copy number changes in genes involved in the PIK3/AKT signalling pathway.

In this study patients presented a high level of DNA copy number changes in genes involved in the *PIK3/AKT* pathway as depicted in Table 3.4 The most frequent changes were observed in the *PIK3CA* and *PTEN* genes. *PIK3CA* presented high copy number in 36.8% of the cases. *PTEN* was observed with low copy number in 47.5% of the cases.

Table 3.4 : DNA copy number assay for *AKT1*, *AKT2* and *PTEN* genes in the non-TNBC.

Gene	**Normal copy number n (%)**	**High copy number n (%)**	**Low copy number n (%)**
ATK1	22 (57.9%)	7 (18.4%)	9 (23.7%)
ATK2	25 (65.8%)	9 (23.7%)	4 (10.5%)
PIK3CA	14 (36.8%)	14 (36.8%)	10 (26.3%)
PTEN	15 (39.5%)	5 (13.1%)	18 (47.4%)
FGFR	20 (52.6%)	7 (18.4%)	11 (28.9%)

3.4.2 DNA copy number assay for *AKT1*, *AKT2* and *PTEN* genes for patients with breast tumor (TNBC and non-TNBC) from the South African population

Patients with breast tumor (TNBC and non-TNBC) from the South African population present a high level of DNA copy number changes in genes involved in the *PIK3/AKT* pathway as indicated in Figure 3.4 and Figure 3.5.

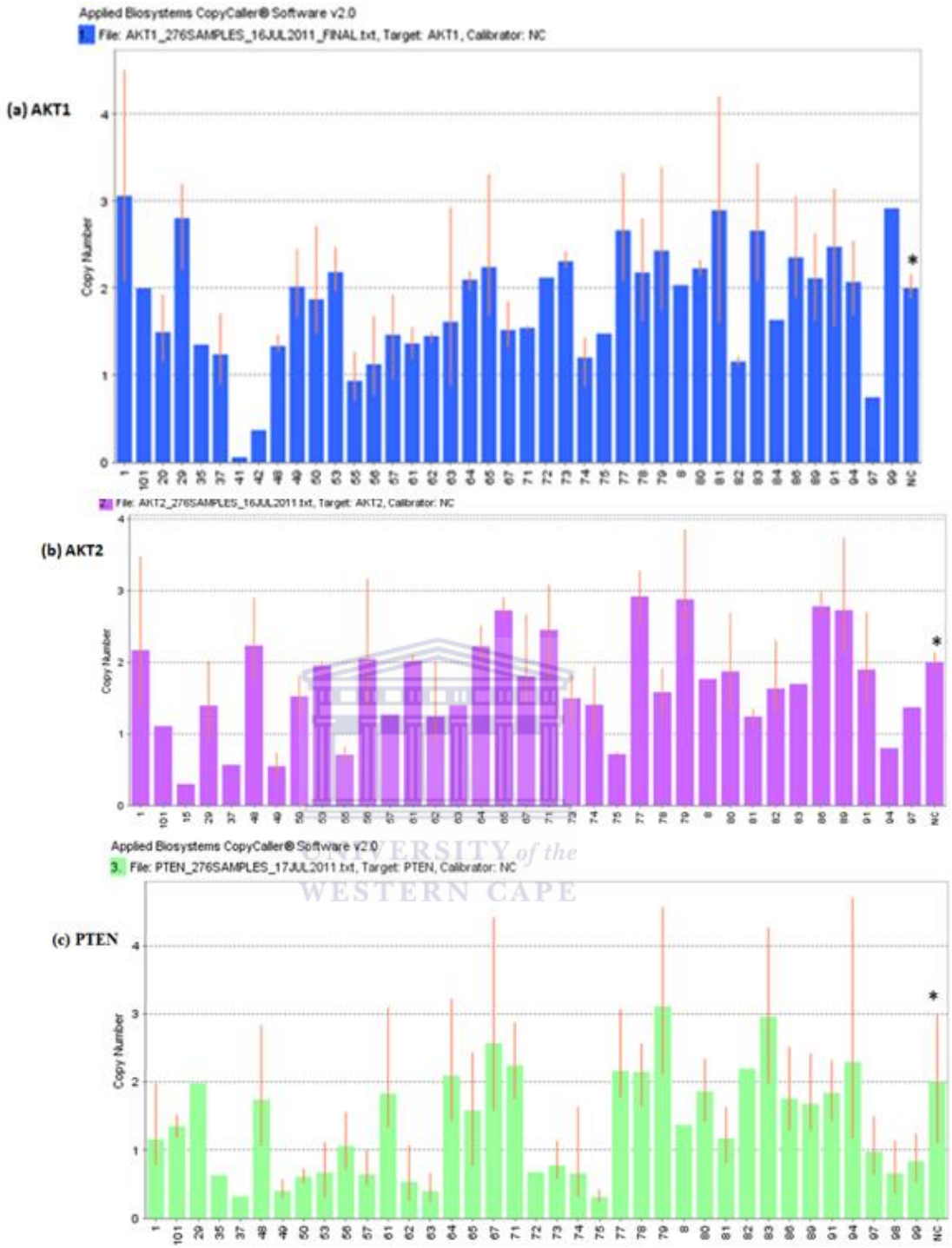


Figure 3.4: Represents DNA copy number assay for *AKT1*, *AKT2* and *PTEN* genes in non- TNBC from South African patients. *indicates a normal control. The y axis represents the DNA copy number and the x axis represents the number of a non TNBC sample from the 101 samples:

(a) For the ***AKT1*** gene samples 1, 101, 20, 29, 35, 37, 41,42, 48, 49, 50, 53, 55, 56, 57, 61, 62, 63, 64. 65, 67, 71, 72, 73, 74, 75, 77, 78 79, 8, 81, 82, 83, 84, 86, 89, 91, 94, 97, 99, NC.

(b) For the ***AKT2*** gene samples: 1, 101, 15, 73, 43, 49, 50, 53, 55, 56, 57, 61, 62, 63, 64, 65, 67, 71, 73, 74, 75, 77, 78, 79, 8, 80, 81, 82, 83, 86, 89, 91, 94, 97, NC

(c) For the ***PTEN*** gene samples: 1, 101, 29, 35, 37, 48, 49, 50, 53, 56, 57, 61, 62, 63, 64, 64, 67, 71, 72, 73, 74, 75, 77, 78, 79, 8, 80, 81, 82, 83, 86, 89, 91, 94, 97, 98, 99, NC.

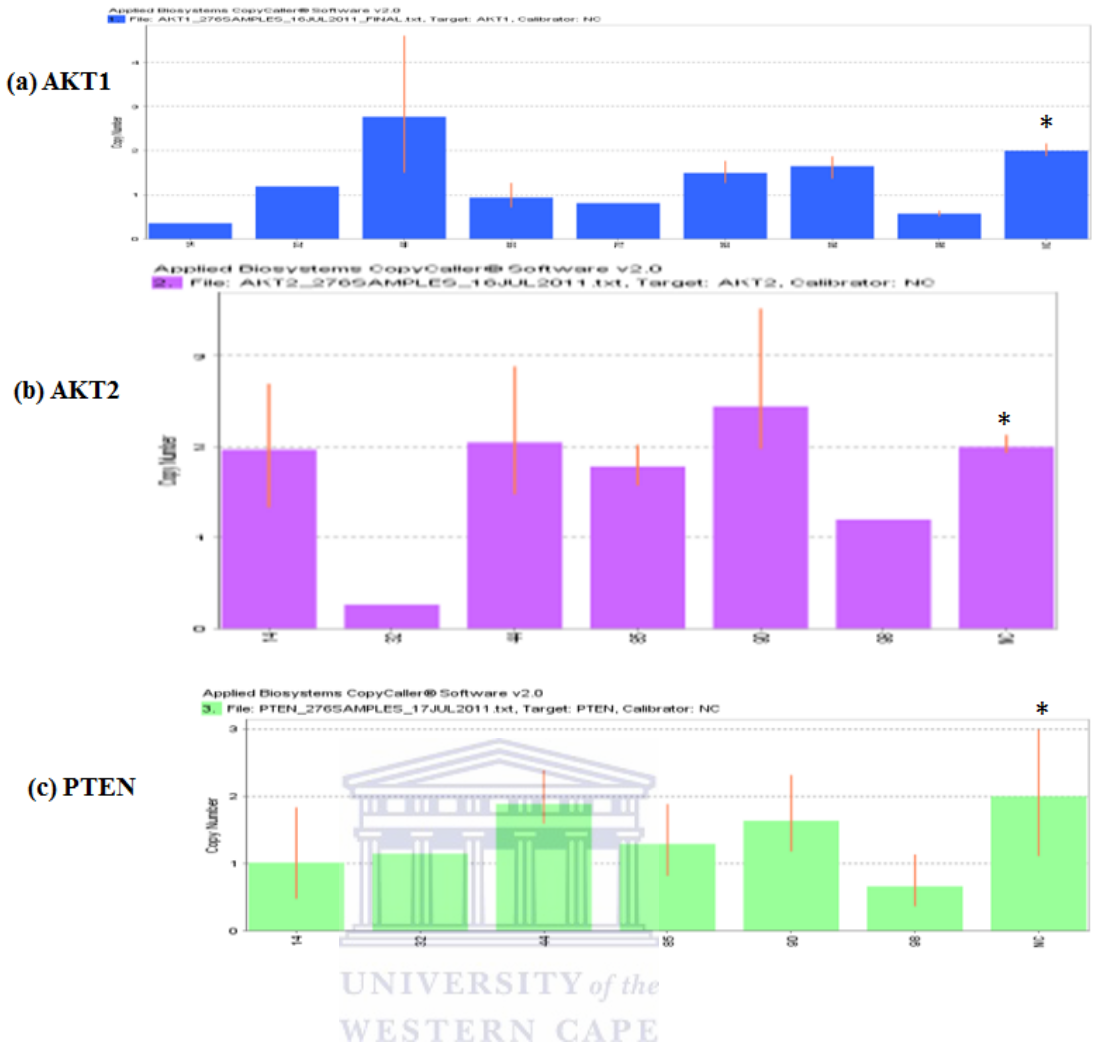


Figure 3.5: DNA copy number assay for (a) *AKT1*, (b) *AKT2* and (c) *PTEN* in the TNBC from South African patients. (*) indicates normal control (NC) . The y axis represents the DNA copy number and the x axis represents the number of TNBC samples from the 101 samples:

(a) For the *AKT1* gene samples: 14, 32, 44, 55, 71, 85, 90, 98 and NC

(b) For the *AKT2* gene samples: 14, 32, 44, 85, 90, 98, NC

(c) For the *PTEN* gene samples: 14, 32, 44, 85, 90, 98, NC

CHAPTER 4

DISCUSSION

The emerging fact that BC has a heterogenous biological behaviour has warranted extensive research. For the past two decades, researchers have tried to stratify BC in order to find improved means for early diagnosis and ultimately better therapeutic approach. As a result of genomic instability, cancer cells have been found to accumulate both random and causal alterations at multiple levels, ranging from point mutations to whole chromosome aberrations (Huang, Wei et al. 2004). Examples of DNA copy number changes include, but are not limited to, loss of heterozygosity (LOH) and homozygous deletions, which may result in the loss of tumor suppressor genes, and gene amplification events, which can activate cellular proto-oncogenes (Huang, Wei et al. 2004).

Like other solid cancers, breast cancer presents genomic instability (Li, Wang et al. 2009). The existing concept is that frequently occurring regions of DNA amplification commonly harbor oncogenes, whereas regions of recurrent deletion are known to harbor tumor suppressor genes (Li, Wang et al. 2009). Classical cytogenetic methods have been used to detect such copy number changes in tumors, which have deepened our understanding of the genomic hallmarks of breast cancer (Li, Wang et al. 2009). The main objective of this study was to find markers for the detection of genomic instability in South African women. The further aim was to be able to correlate these genomic changes with the age, receptor status and staging, to gain an insight into the prognostic significance of these markers.

The specimen chosen for this study were primary tumors obtained from the archives of National Laboratory Health Sciences (NLHS) after obtaining ethics approval. Formalin-fixed, paraffin-embedded (FFPE) tissues are one of the popular sources of diagnostic materials, the easiest to store and transport. They are often used as the source of nucleic acids for retrospective molecular analyses based on DNA amplification by polymerase chain reaction (PCR). A study has shown that the LOH is better detected in tumor tissues than haematological specimens like blood and bone marrow (Schwarzenbach, Muller et al. 2007). However, it is known that nucleic acids from paraffin embedded tissues are much worse templates than those recovered from fresh tissues (Bielawski, Zaczek et al. 2001) and is mainly due to the low yield of genetic material and the difficulty in getting intact DNA.

The low yield of DNA from our archival samples was one of the major obstacles in this study. This has been reported in other studies (Al-Mulla 2011). One of the future approaches to achieve high throughput for LOH mapping is through the use of Capillary array electrophoresis and laser-assisted microdissection (Ellsworth, Shriver et al. 2003). Mutiple sections were explored in order to ensure the generation of intact DNA material.

The choice of control was problematic as we initially intended to use blood samples as controls, however, the ethics approval committee has objected to that and we were only left with the use of normal breast tissue as a control. A study evaluated LOH in normal breast epithelial tissues and benign breast lesions around the tumor and found them to harbour LOH (Cavalli, Singh et al. 2004). To overcome this, we have asked the pathologist to provide samples of normal breast tissue away from the perimeter of the tumor but blood samples remains as more suitable control.

Clinically advanced tumor stage and tumor grade are recognized prognostic factors in breast carcinoma (Henson, Ries et al. 1991; Page 1997). Gupta et.al (2009) found that all the tumors having LOH, were of high grade (grade II and III). These findings suggest that the accumulation of allelic losses might play an important role in tumor progression(Gupta, Joshi et al. 2009). In contrast, there is no significant difference in LOH frequency with respect to age, tumor stage, menopausal status, chemotherapy or lymph node metastasis (Hirano, Emi et al. 2001).

In this study, the association between LOH and stage is not statistically significant. There was no significance difference in early stage (15.8%) as compared to late stage (7.9%) in determining the LOH effect.

To detect genomic instability, two approaches were used; the first was to attempt to detect allelic instability or LOH at the *IRF-1* region of chromosome 5.

The loss of heterozygosity indicates that at one of the constitutional loci, a normally functioning allele was lost (Skotheim. 2001; Miller, Wang et al. 2003). This loss is commonly found in different types of cancers, including breast cancer, cervical cancer and squamous cell carcinoma (Huettner, Gerhard et al. 1998; Miller, Wang et al. 2003). In most types of cancer this loss can be mapped, should it occur in a non-random fashion, and thereby implicating specific tumor suppressor genes in the progression of a cancerous state (Skotheim. 2001). Total knockout of a tumor suppressor gene can occur, whereby one allele has been mutated or imprinted, and this is enhanced by the loss of heterozygosity (Skotheim. 2001). To detect for differences in two alleles at a locus, the intensity ration can be measured (Skotheim. 2001; Miller, Wang et al. 2003). A skewed intensity ratio is known as allelic imbalance and this indicates the possible loss of an allele due to masking, heterogeneity or non-clonal loss (Skotheim. 2001; Miller, Wang et al. 2003).

The choice of the IRF1 gene has been based on reports of loss of heterozygosity and deletions associated with chrosome 5 in TNBC and BRCA1 cases (Wang, Lin et al. 2004; Bergamaschi, Kim et al. 2006). In addition, other researchers have reported on its characteristics as a promising candidate marker for the detection of loss of heterozygosity in primary breast tumors (Cavalli, Riggins et al. 2010). It was also documented that IRF1 has a remarkable functional diversity and controls the transcription of genes involved in mediating antiviral, immunomodulatory, and antiproliferative effects (Tamura, Tailor et al. 2005). It was postulated that *IRF-1* can act as a potential tumor suppressor gene through the regulation of oncogenesis and elimination of precancerous cells through apoptosis induced by DNA damage or other stimuli (Tamura, Ishihara et al. 1995).

In this study allelic loss of *IRF-1* was detected in 23/46 (50%) breast tissue specimens. This is slightly lower than what is reported by another study on American women where 52 women were investigated for LOH at *IRF-1* locus. 37 cases were informative and had presented a 32% (12 /37) LOH at *IRF-1* (Cavalli 2009).

Although studies of mammalian cells consistently reveal an increased frequency of mutations with advanced age, the experimental data to date do not clearly demonstrate any age-dependent change in the rate of LOH in vivo. Several studies demonstrate an overall functional decline in different types of DNA repair with age. Perhaps the most interesting finding in the studies of age-induced LOH, is the evidence for a change in the mechanism of repair (Carr and Gottschling 2008).

In this study, there was no statistical significance (P=1.0000) between age (≤50 years or >50 years) with the presence or absence of LOH. In summary, our understanding of the association between age-induced LOH remains incomplete.

DNA copy number variations (CNVs) are seen to be an important component of genetic deviation, which is known to affect a greater fraction of the genome than single nucleotide polymorphisms (Shlien and Malkin 2009). The characterization of widespread CNVs has provided insight into their role in susceptibility to an extensive range of diseases, and has been found that somatic CNVs can be used to identify regions of the genome involved in disease phenotypes (Shlien and Malkin 2009). Ultimately, changes in DNA copy number can be associated with many human cancers (Huang, Madireddi et al. 2001; Huang, Wei et al. 2004; Wang, Lin et al. 2004). A very interesting recent study compared CNVs between FFPE samples from 104 African (Malian) women who had undergone mastectomy and another cohort of patients from USA and found significant differences in genomic profiling (GNVs) in six chromosomal regions (Madanikia, Bergner et al. 2012).

In this study, our second approach was to determine the DNA copy number changes of specific genes involved in the *PIK3/AKT* signaling pathway in our samples. Genes associated with *PIK3* pathway are known to be expressed in 50% of BC (Samuels and Ericson 2006). *PI3K* signaling is activated in human cancers via several different mechanisms (Samuels and Velculescu 2004). Increased *PI3K* signaling is often due to direct mutational activation or amplification of genes encoding key components of the *PI3K* pathway such as *PIK3CA* and *AKT1*, or loss of *PTEN* (Bader, Kang et al. 2005; Samuels and Ericson 2006).

The most common genetic alteration of the *PI3K* signalling pathway found in human cancer is inactivation of the *PTEN* tumor suppressor gene. Inactivation of *PTEN* leads to loss of its lipid phosphatase activity, causing accumulation of PIP3 (Haas-Kogan, Shalev et al. 1998; Myers, Pass et al. 1998).

In this study patients with breast tumor (TNBC and non-TNBC population) presented a high level of DNA copy number changes in genes involved in the *PIK3/AKT* pathway. The most frequent changes were observed in the *PIK3CA* and *PTEN* genes. *PIK3CA* presented high copy number in 36.8% of the cases. *PTEN* was observed with low copy number in 47.5% of the cases.

For *ATK1* gene, 22 (57.9%) cases had normal copy number, 7 (18.4%) cases presented a high copy number and 9 (23.7%) cases presented a low copy number. For the *ATK2* gene 25 (65.8%) cases presented a normal copy number, 9 (23.7%) cases presented a high copy number and 4 (10.5%) cases presented a low copy number. For the *PIK3CA* gene 14 (36.8%) cases presented a normal copy number, 14 (36.8%) cases presented a high copy number and 10 (26.3) cases presented a low copy number. For the *PTEN* gene 15 (39.5%) presented a normal copy number and 5 (13.1%) high copy number and 18 (47.4%) a low copy number. For *FGFR* gene 20 (52.6%) presented a normal copy number, 7 (18.45%) a high copy number and 11 (28.9%) a low copy number. Furthermore, there was a significant difference in our indicators when compared with a Brazilian cohort (Unpublished data).

The Copy Caller Software (Applied Biosystems, CA) results revealed alterations in the DNA copy number in the samples analyzed. These changes were very significant and was at a higher level than the control group (please refer to the

results section) indicating the suitability of these genes for breast cancer detection in South African patients.

The further aim of this study was to correlate our genomic findings with the clinical follow-up data obtained for our candidate genes in order to establish whether these genes have prognostic significance. One of the concepts explored was the concept of TNBC, defined by a lack of expression of both estrogen (ER) and progesterone (PgR) receptors as well as human epidermal growth factor receptor 2 (HER2) (Ovcaricek, Frkovic et al. 2011). Since this phenotype is common among women of African descent and know to have poor prognosis, we were keen to find out the percentage of TNBC in our cases. In this study, 27 cases (26.73%) out of 101 were triple negative (TBNC). The ER/PR pharmDx™ Scoring system interpretation manual (Allred scoring guidelines) was used to analyse the data. However, a bigger sample size that is not ethnically diverse might give a better reflection of the actual predominance of TNBC in SA populations.

CONCLUSION AND FUTURE OUTLOOK

Genetic instability is an early indication of malignant changes. Two approaches were adopted in this study to detect genomic instability: Loss of heterozygosity and DNA copy number variations.

Our data analysis indicated that 46 cases (45.5%) out of 101 cases were informative for the *IRF-1* dinucleotide marker used for LOH analysis. LOH was detected in 23 of these informative cases (23/46; 50%) which shows that *IRF-1* is a good marker for early BC detection in primary tumors. Furthermore, of the genes involved in the *PIK3* pathway, the most frequent changes were observed in the *PIK3CA* and *PTEN* genes. *PIK3CA* presented high copy number in 36.8% of the cases. *PTEN* was observed with low copy number in 47.5% of the cases when compared with the control indicating their validity as BC markers.

We believe that future research should focus on other means of detecting genomic instability. The actual percentages of TNBC patients still need to be verified.

CHAPTER 5

REFERENCES

Aapro, M. and H. Wildiers (2012). "Triple-negative breast cancer in the older population." Ann Oncol **23 Suppl 6**: vi52-55.

Al-Mulla, F. (2011). "Microarray-based CGH and copy number analysis of FFPE samples." Methods Mol Biol **724**: 131-145.

Ameyaw, M. M., N. Thornton, et al. (2000). "Re: population-based, case-control study of HER2 genetic polymorphism and breast cancer risk." J Natl Cancer Inst **92**(23): 1947.

Amir, H., M. R. Azizi, et al. (1997). "TNM classification and breast cancer in an African population: a descriptive study." Cent Afr J Med **43**(12): 357-359.

Amir, H., J. N. Kitinya, et al. (1994). "A comparative study of carcinoma of the breast in an African population." East Afr Med J **71**(4): 215-218.

Anders, C. and L. A. Carey (2008). "Understanding and treating triple-negative breast cancer." Oncology (Williston Park) **22**(11): 1233-1239; discussion 1239-1240, 1243.

August, D. A., T. Rea, et al. (1994). "Age-related differences in breast cancer treatment." Ann Surg Oncol **1**(1): 45-52.

Bader, A. G., S. Kang, et al. (2005). "Oncogenic PI3K deregulates transcription and translation." Nat Rev Cancer **5**(12): 921-929.

Ballestrero, A., D. A. Coviello, et al. (2001). "Reverse-transcriptase polymerase chain reaction of the maspin gene in the detection of bone marrow breast carcinoma cell contamination." Cancer **92**(8): 2030-2035.

Basro, S. and J. P. Apffelstaedt (2010). "Breast cancer in young women in a limited-resource environment." World J Surg **34**(7): 1427-1433.

Bergamaschi, A., Y. H. Kim, et al. (2006). "Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer." Genes Chromosomes Cancer **45**(11): 1033-1040.

Berry, D. A., K. A. Cronin, et al. (2005). "Effect of screening and adjuvant therapy on mortality from breast cancer." N Engl J Med **353**(17): 1784-1792.

Bielawski, K., A. Zaczek, et al. (2001). "The suitability of DNA extracted from formalin-fixed, paraffin-embedded tissues for double differential polymerase chain reaction analysis." Int J Mol Med **8**(5): 573-578.

Bouker, K. B., T. C. Skaar, et al. (2005). "Interferon regulatory factor-1 (IRF-1) exhibits tumor suppressor activities in breast cancer associated with caspase activation and induction of apoptosis." Carcinogenesis **26**(9): 1527-1535.

Boyle, P. (2012). "Triple-negative breast cancer: epidemiological considerations and recommendations." Ann Oncol **23 Suppl 6**: vi7-12.

Boyle, P., M. E. Leon, et al. (2003). "Cancer control in women. Update 2003." Int J Gynaecol Obstet **83 Suppl 1**: 179-202.

Bradshaw, E. and J. S. Harington (1982). "A comparison of cancer mortality rates in South Africa with those in other countries." S Afr Med J **61**(25): 943-946.

Braun, S., K. Pantel, et al. (2000). "Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer." N Engl J Med **342**(8): 525-533.

Brenner, A. J. and C. M. Aldaz (1997). "The genetics of sporadic breast cancer." Prog Clin Biol Res **396**: 63-82.

Brown, L. W. and R. D. Williams (1994). "Culturally sensitive breast cancer screening programs for older black women." Nurse Pract **19**(3): 21, 25-26, 31 passim.

Carr, L. L. and D. E. Gottschling (2008). "Does age influence loss of heterozygosity?" Exp Gerontol **43**(3): 123-129.

Carracedo, A., L. Ma, et al. (2008). "Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer." J Clin Invest **118**(9): 3065-3074.

Cavalli, L. R. (2009). "Molecular markers of breast axillary lymph node metastasis." Expert Rev Mol Diagn. **9**(5): 441-454.

Cavalli, L. R., R. B. Riggins, et al. (2010). "Frequent loss of heterozygosity at the interferon regulatory factor-1 gene locus in breast cancer." Breast Cancer Res Treat **121**(1): 227-231.

Cavalli, L. R., B. Singh, et al. (2004). "Loss of heterozygosity in normal breast epithelial tissue and benign breast lesions in BRCA1/2 carriers with breast cancer." Cancer Genet Cytogenet **149**(1): 38-43.

Cavenee, W. K. (1989). "Tumor progression stage: specific losses of heterozygosity." Princess Takamatsu Symp **20**: 33-42.

Charef-Hamza, S., M. Trimeche, et al. (2005). "Loss of heterozygosity at the BRCA1 locus in Tunisian women with sporadic breast cancer." Cancer Lett **224**(2): 185-191.

Chavez, K. J., S. V. Garimella, et al. (2010). "Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer." Breast Dis **32**(1-2): 35-48.

Chu, K. C., R. E. Tarone, et al. (1996). "Recent trends in U.S. breast cancer incidence, survival, and mortality rates." J Natl Cancer Inst **88**(21): 1571-1579.

Clegg-Lamptey, J. N., J. C. Dakubo, et al. (2009). "Psychosocial aspects of breast cancer treatment in Accra, Ghana." East Afr Med J **86**(7): 348-353.

Cserni, G. (2004). "Surgical pathological staging of breast cancer by sentinel lymph node biopsy with special emphasis on the histological work-up of axillary sentinel lymph nodes." Breast Cancer. **11**(3): 242-249; discussion 264-246.

Cullen, R., T. M. Maguire, et al. (2001). "Studies on oestrogen receptor-alpha and -beta mRNA in breast cancer." Eur J Cancer **37**(9): 1118-1122.

Dawood, S. (2010). "Triple-negative breast cancer: epidemiology and management options." Drugs **70**(17): 2247-2258.

Dillon, R. L., D. E. White, et al. (2007). "The phosphatidyl inositol 3-kinase signaling network: implications for human breast cancer." Oncogene **26**(9): 1338-1345.

Domschke, C., Y. Ge, et al. (2013). "Long-term survival after adoptive bone marrow T cell therapy of advanced metastasized breast cancer: follow-up analysis of a clinical pilot trial." Cancer Immunol Immunother **62**(6): 1053-1060.

Dubois, J. B. (1990). "[Treatment of tumors of the breast less than 3 cm in diameter]." Rev Prat **40**(10): 890-893.

El Khouli, R. H., K. J. Macura, et al. (2009). "Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast." J Magn Reson Imaging **30**(5): 999-1004.

Ellis, M. J., L. Lin, et al. (2010). "Phosphatidyl-inositol-3-kinase alpha catalytic subunit mutation and response to neoadjuvant endocrine therapy for estrogen receptor positive breast cancer." Breast Cancer Res Treat **119**(2): 379-390.

Ellis, M. J., W. R. Miller, et al. (2009). "Aromatase expression and outcomes in the P024 neoadjuvant endocrine therapy trial." Breast Cancer Res Treat **116**(2): 371-378.

Ellsworth, D. L., C. D. Shriver, et al. (2003). "Laser capture microdissection of paraffin-embedded tissues." Biotechniques **34**(1): 42-44, 46.

Fisher, B. (1999). "From Halsted to prevention and beyond: advances in the management of breast cancer during the twentieth century." Eur J Cancer. **35**(14): 1963-1973.

Fisher, B., M. Bauer, et al. (1983). "Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update." Cancer **52**(9): 1551-1557.

Fisher, C. S., D. J. Cole, et al. (2010). "Molecular detection of micrometastatic breast cancer in histopathology-negative axillary lymph nodes fails to predict breast cancer recurrence: a final analysis of a prospective multi-institutional cohort study." Ann Surg Oncol **17 Suppl 3**: 312-320.

Fodor, J., E. Mozsa, et al. (2005). "[Local relapse in young (< or = 40 years) women with breast cancer after mastectomy or breast conserving surgery: 15-year results]." Magy Onkol. **49**(3): 203, 205-208 Epub 2005 Oct 2025.

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

Fujii, H., Y. Ajioka, et al. (2002). "Loss of heterozygosity in the clonal evolution of flat colorectal neoplasms." J Pathol **197**(3): 298-306.

Giuffre, G., V. Adamo, et al. (2011). "Hematopoietic progenitor cells (HPCs) in node-negative invasive breast carcinomas: Immunohistochemical analysis and clinico-pathological correlations." Pathol Res Pract **207**(8): 487-491.

Giuliano, A. E., K. K. Hunt, et al. (2011). "Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial." JAMA **305**(6): 569-575.

Gluz, O., C. Liedtke, et al. (2009). "Triple-negative breast cancer--current status and future directions." Ann Oncol **20**(12): 1913-1927.

Goedecke, J. H., J. Evans, et al. (2011). "Reduced gluteal expression of adipogenic and lipogenic genes in Black South African women is associated with obesity-related insulin resistance." J Clin Endocrinol Metab **96**(12): E2029-2033.

Guenther, J. M., N. M. Hansen, et al. (2003). "Axillary dissection is not required for all patients with breast cancer and positive sentinel nodes." Arch Surg **138**(1): 52-56.

Gupta, S., K. Joshi, et al. (2009). "High frequency of loss of allelic integrity at Wilms' tumor suppressor gene-1 locus in advanced breast tumors associated with aggressiveness of the tumor." Indian J Cancer **46**(4): 303-310.

Haas-Kogan, D., N. Shalev, et al. (1998). "Protein kinase B (PKB/Akt) activity is elevated in glioblastoma cells due to mutation of the tumor suppressor PTEN/MMAC." Curr Biol **8**(21): 1195-1198.

Haffty, B. G., C. McKhann, et al. (1993). "Breast conservation therapy without axillary dissection. A rational treatment strategy in selected patients." Arch Surg **128**(12): 1315-1319; discussion 1319.

Halsted, W. S. (1907). "The results of radical operations for the cure of carcinoma of the breast." Ann Surg **46**(1).

Hassan, I., E. E. Onukak, et al. (1992). "Breast cancer in Zaria, Nigeria." J R Coll Surg Edinb **37**(3): 159-161.

Helman, P. (1977). "Whither breast cancer? Report on the inaugural meeting of the National Breast Cancer Group." S Afr Med J **52**(18): 711-713.

Henderson, B. E., R. K. Ross, et al. (1982). "Endogenous hormones as a major factor in human cancer." Cancer Res **42**(8): 3232-3239.

Henson, D. E., L. Ries, et al. (1991). "Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index." Cancer **68**(10): 2142-2149.

Hermanek, P. (1999). "Prognostic factor research in oncology." J Clin Epidemiol **52**(4): 371-374.

Hilakivi-Clarke, L., C. Wang, et al. (2004). "Nutritional modulation of the cell cycle and breast cancer." Endocr Relat Cancer **11**(4): 603-622.

Hirano, A., M. Emi, et al. (2001). "Allelic losses of loci at 3p25.1, 8p22, 13q12, 17p13.3, and 22q13 correlate with postoperative recurrence in breast cancer." Clin Cancer Res **7**(4): 876-882.

Holmes, D. (2011). "PI3K pathway inhibitors approach junction." Nat Rev Drug Discov **10**(8): 563-564.

Huang, E. Y., M. T. Madireddi, et al. (2001). "Genomic structure, chromosomal localization and expression profile of a novel melanoma differentiation associated (mda-7) gene with cancer specific growth suppressing and apoptosis inducing properties." Oncogene **20**(48): 7051-7063.

Huang, J., W. Wei, et al. (2004). "Whole genome DNA copy number changes identified by high density oligonucleotide arrays." Hum Genomics **1**(4): 287-299.

Huettner, P. C., D. S. Gerhard, et al. (1998). "Loss of heterozygosity in clinical stage IB cervical carcinoma: relationship with clinical and histopathologic features." Hum Pathol **29**(4): 364-370.

Irvin, W. J., Jr. and L. A. Carey (2008). "What is triple-negative breast cancer?" Eur J Cancer **44**(18): 2799-2805.

Isaacs, C., L. R. Cavalli, et al. (2004). "Detection of LOH and mitochondrial DNA alterations in ductal lavage and nipple aspirate fluids from hngh-risk patients." Breast Cancer Res Treat **84**(2): 99-105.

Johnson, P. J. and Y. M. Lo (2002). "Plasma nucleic acids in the diagnosis and management of malignant disease." Clin Chem **48**(8): 1186-1193.

Kim, P. K., M. Armstrong, et al. (2004). "IRF-1 expression induces apoptosis and inhibits tumor growth in mouse mammary cancer cells in vitro and in vivo." Oncogene **23**(5): 1125-1135.

Kiricuta, I. C. (2000). "Sentinel node concept in breast cancer." Strahlenther Onkol **176**(7): 307-314.

Lasko, D., W. Cavenee, et al. (1991). "Loss of constitutional heterozygosity in human cancer." Annu Rev Genet **25**: 281-314.

Lehmann, B. D., J. A. Bauer, et al. (2011). "Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies." J Clin Invest **121**(7): 2750-2767.

Li, J., K. Wang, et al. (2009). "DNA Copy Number Aberrations in Breast Cancer by Array Comparative Genomic Hybridization." Genomics, Proteomics & Bioinformatics **7**(1-2): 13-24.

Lin, Y. S., X. Y. Zhong, et al. (2000). "Detection of tumor cells in leukapheresis products from patients with breast cancer using immunocytochemical staining method." Arch Gynecol Obstet **263**(3): 119-125.

Logue, J. S. and D. K. Morrison (2012). "Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy." Genes Dev **26**(7): 641-650.

Lurie, N. and R. Parker (2007). "Editorial: moving health literacy from the individual to the community." Am J Health Behav **31 Suppl 1**: S6-7.

Madanikia, S. A., A. Bergner, et al. (2012). "Increased risk of breast cancer in women with NF1." Am J Med Genet A **158A**(12): 3056-3060.

McCurtis Witherspoon, K. and A. Williams Richardson (2006). "Sisters in support together against substances (SISTAS): an alcohol abuse prevention group for Black women." J Ethn Subst Abuse **5**(3): 49-60.

McNeely, M. L., J. M. Binkley, et al. (2012). "A prospective model of care for breast cancer rehabilitation: postoperative and postreconstructive issues." Cancer **118**(8 Suppl): 2226-2236.

Meng, Y., S. Ward, et al. (2011). "Cost-effectiveness of MRI and PET imaging for the evaluation of axillary lymph node metastases in early stage breast cancer." Eur J Surg Oncol **37**(1): 40-46.

Miller, B. J., D. Wang, et al. (2003). "Pooled analysis of loss of heterozygosity in breast cancer: a genome scan provides comparative evidence for multiple tumor suppressors and identifies novel candidate regions." Am J Hum Genet **73**(4): 748-767.

Muguti, G. I. (1993). "Experience with breast cancer in Zimbabwe." J R Coll Surg Edinb **38**(2): 75-78.

Murthy, S. K., L. M. DiFrancesco, et al. (2002). "Loss of heterozygosity associated with uniparental disomy in breast carcinoma." Mod Pathol **15**(12): 1241-1250.

Myers, M. P., I. Pass, et al. (1998). "The lipid phosphatase activity of PTEN is critical for its tumor supressor function." Proc Natl Acad Sci U S A **95**(23): 13513-13518.

Nathanson, K. L., R. Wooster, et al. (2001). "Breast cancer genetics: what we know and what we need." Nat Med **7**(5): 552-556.

Nathanson, S. D. (1999). "Will the true sentinel node please stand?" Ann Surg Oncol **6**(5): 514-516.

Nishimura, R., T. Osako, et al. (2010). "Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer." Exp Ther Med **1**(5): 747-754.

O'Higgins, N., D. A. Linos, et al. (1998). "European guidelines for quality assurance in the surgical management of mammographically detected lesions." Eur J Surg Oncol **24**(2): 96-98.

Ovcaricek, T., S. G. Frkovic, et al. (2011). "Triple negative breast cancer - prognostic factors and survival." Radiol Oncol **45**(1): 46-52.

Page, D. L. (1997). "Prognostic indicators in breast cancer and who needs them." Anat Pathol **2**: 35-52.

Parkin, D. M., P. Pisani, et al. (1999). "Global cancer statistics." CA Cancer J Clin **49**(1): 33-64, 31.

Parkin, D. M., F. Sitas, et al. (2008). "Part I: Cancer in Indigenous Africans--burden, distribution, and trends." Lancet Oncol **9**(7): 683-692.

Perou, C. M. (2010). "Molecular stratification of triple-negative breast cancers." Oncologist **15 Suppl 5**: 39-48.

Petrek, J. A., R. T. Senie, et al. (2001). "Lymphedema in a cohort of breast carcinoma survivors 20 years after diagnosis." Cancer. **92**(6): 1368-1377.

Pillay, A. L. (2002). "Rural and urban South African women's awareness of cancers of the breast and cervix." Ethn Health **7**(2): 103-114.

Reeves, M. D., T. M. Yawitch, et al. (2004). "BRCA1 mutations in South African breast and/or ovarian cancer families: evidence of a novel founder mutation in Afrikaner families." Int J Cancer **110**(5): 677-682.

Richard, D. (2009). "Gray's Anatomy for Students."

Rosner, D. and W. W. Lane (1993). "Predicting recurrence in axillary-node negative breast cancer patients." Breast Cancer Res Treat. **25**(2): 127-139.

Rottenberg, S., J. E. Jaspers, et al. (2008). "High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs." Proc Natl Acad Sci U S A **105**(44): 17079-17084.

Samuels, Y. and K. Ericson (2006). "Oncogenic PI3K and its role in cancer." Curr Opin Oncol **18**(1): 77-82.

Samuels, Y. and V. E. Velculescu (2004). "Oncogenic mutations of PIK3CA in human cancers." Cell Cycle **3**(10): 1221-1224.

Schwarzenbach, H., V. Muller, et al. (2007). "A critical evaluation of loss of heterozygosity detected in tumor tissues, blood serum and bone marrow plasma from patients with breast cancer." Breast Cancer Res **9**(5): R66.

Shlien, A. and D. Malkin (2009). "Copy number variations and cancer." Genome Med **1**(6): 62.

Silva, J. M., G. Dominguez, et al. (1999). "Presence of tumor DNA in plasma of breast cancer patients: clinicopathological correlations

Loss of heterozygosity in BRCA1 and BRCA2 markers and high-grade malignancy in breast cancer." Cancer Res **59**(13): 3251-3256.

Sjogren, S., M. Inganas, et al. (1996). "The p53 gene in breast cancer: prognostic value of complementary DNA sequencing versus immunohistochemistry." J Natl Cancer Inst **88**(3-4): 173-182.

Skotheim., e. a. (2001). "Evaluation of loss of heterozygosity/allelic imbalance scoring in tumor DNA." Cancer genetics and cytogenetics **127**: 64-70.

Sobin, L. H., P. Hermanek, et al. (1988). "TNM classification of malignant tumors. A comparison between the new (1987) and the old editions." Cancer **61**(11): 2310-2314.

Somdyala, N. I., D. Bradshaw, et al. (2010). "Cancer incidence in a rural population of South Africa, 1998-2002." Int J Cancer **127**(10): 2420-2429.

Sorlie, T., C. M. Perou, et al. (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A **98**(19): 10869-10874.

St-Germain, M. E., V. Gagnon, et al. (2004). "Regulation of COX-2 protein expression by Akt in endometrial cancer cells is mediated through NF-kappaB/IkappaB pathway." Mol Cancer **3**: 7.

Stevens, A., J. S. Lowe, et al. (2002). "Breast in Wheather's Basic Histology." 216-223.

Stevens, K. N., Z. Fredericksen, et al. (2012). "19p13.1 is a triple-negative-specific breast cancer susceptibility locus." Cancer Res **72**(7): 1795-1803.

Suter, R. and J. A. Marcum (2007). "The molecular genetics of breast cancer and targeted therapy." Biologics **1**(3): 241-258.

Tabar, L., M. F. Yen, et al. (2003). "Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening." Lancet **361**(9367): 1405-1410.

Tamura, T., M. Ishihara, et al. (1995). "An IRF-1-dependent pathway of DNA damage-induced apoptosis in mitogen-activated T lymphocytes." Nature **376**(6541): 596-599.

Tamura, T., P. Tailor, et al. (2005). "IFN regulatory factor-4 and -8 govern dendritic cell subset development and their functional diversity." J Immunol **174**(5): 2573-2581.

Thomas, D. B., R. A. Carter, et al. (2002). "Risk of subsequent breast cancer in relation to characteristics of screening mammograms from women less than 50 years of age." Cancer Epidemiol Biomarkers Prev **11**(6): 565-571.

van Bogaert, L. J. (2013). "Breast cancer molecular subtypes as identified by immunohistochemistry in South African black women." Breast J **19**(2): 210-211.

van der Merwe, N. C., N. Hamel, et al. (2012). "A founder BRCA2 mutation in non-Afrikaner breast cancer patients of the Western Cape of South Africa." Clin Genet **81**(2): 179-184.

Vorobiof, D. A., F. Sitas, et al. (2001). "Breast cancer incidence in South Africa." J Clin Oncol. **19**(18 Suppl): 125S-127S.

Walker, A. R., F. I. Adam, et al. (2004). "Breast cancer in black African women: a changing situation." J R Soc Promot Health **124**(2): 81-85.

Walker, A. R., B. F. Walker, et al. (1984). "Age of menopause in black women in South Africa." Br J Obstet Gynaecol **91**(8): 797-801.

Wang, Q., P. S. Larson, et al. (2003). "Loss of heterozygosity in serial plasma DNA samples during follow-up of women with breast cancer." Int J Cancer **106**(6): 923-929.

Wang, Z. C., M. Lin, et al. (2004). "Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers." Cancer Res **64**(1): 64-71.

Weis, S. M. and D. A. Cheresh (2011). "Tumor angiogenesis: molecular pathways and therapeutic targets." Nat Med **17**(11): 1359-1370.

Wernicke, A. G., R. L. Goodman, et al. (2011). "A 10-year follow-up of treatment outcomes in patients with early stage breast cancer and clinically negative axillary nodes treated with tangential breast irradiation following sentinel lymph node dissection or axillary clearance." Breast Cancer Res Treat **125**(3): 893-902.

Zhang, H., J. Fang, et al. (2010). "Activation of FOXO1 is critical for the anticancer effect of methylseleninic acid in prostate cancer cells." Prostate **70**(12): 1265-1273.

Appendices

Appendix I (a)

DNA quantification table: DNA yield for each sample was quantified using a Thermo Scientific Nanodrop 1000 Spectrophotometer.

Samples 1-101 DNA concentration for tumor tissue

Sample	ng/µl	260/280
1	343.1	2.05
2	527.9	2.25
3	2245.2	2.02
4	1954.0	1.94
5	1442.5	1.86
6	1194.1	1.82
7	1462.9	1.84
8	1237.5	1.72
9	1198.2	1.78
10	1428.9	1.82
11	1461.4	1.82
12	1436.7	1.84
13	1246.0	1.80
14	1284.6	1.81
15	1137.4	1.78
16	1813.6	1.86
17	1534.8	1.82
18	1646.4	1.84
19	1828.1	1.84
20	1367.1	1.79
21	419.5	2.01
22	400.2	1.97
23	432.8	2.03
24	445.2	2.09
25	405.6	2.01
26	557.2	2.15
27	484.7	2.04
28	341.9	1.94

28	772.6	1.81
30	411.6	2.07
31	705.4	1.76
32	464.6	2.04
33	337.6	1.94
34	437.7	2.00
35	496.3	1.99
36	445.5	2.07
37	521.9	2.04
38	425.3	2.03
39	342.7	1.93
40	437.2	1.96
41	54.8	1.91
42	31.7	1.68
43	53.9	1.79
44	64.3	1.73
45	117.0	1.85
46	95.3	1.77
47	108.6	1.81
48	101.1	1.75
49	109.4	1.78
50	47.1	1.64
51	27.2	1.57
52	60.7	1.79
53	17.8	1.50
54	37.6	1.72
55	105.6	1.79
56	20.5	1.46
57	74.5	1.72
58	105.6	1.79
59	86.0	1.77
60	18.1	1.44
61	111.4	1.76
62	136.5	1.76
63	131.8	1.78
64	118.6	1.66
65	168.4	1.82
66	70.3	1.72
67	101.9	1.77

68	137.4	1.76
69	159.4	1.82
70	163.9	1.81
71	48.8	1.74
72	93.0	1.73
73	115.3	1.78
74	158.9	1.79
75	110.7	1.76
76	90.4	1.75
77	108.5	1.72
78	128.2	1.74
79	131.7	1.77
80	157.8	1.76
81	40.7	1.65
82	140.8	1.69
83	345.7	1.72
84	360.2	1.79
85	314.2	1.75
86	217.9	1.70
87	228.0	1.69
88	884.1	1.54
89	219.2	1.63
90	188.2	1.65
91	215.7	1.68
92	202.7	1.65
93	332.1	1.79
94	121.0	1.66
95	102.5	1.65
96	567.8	2.07
97	332.3	1.75
98	418.5	1.84
99	333.3	1.75
100	344.6	1.75
101	247.4	1.68

Appendix I (b)

DNA quantification table: DNA yield for each sample was quantified using a Thermo Scientific Nanodrop 1000 Spectrophotometer.

Samples 1-101 DNA concentration for normal tissue

Sample number	ng/µl	260/280
1	16.8	1.36
2	38.2	1.37
3	26.6	1.51
4	43.4	1.57
5	43.4	1.5
6	28.2	1.39
7	59.5	1.58
8	58.3	1.57
9	40.1	1.47
10	22.9	1.32
11	3.7	0.84
12	16	1.29
13	76.8	1.68
14	200	1.21
15	42.1	1.59
16	38.4	1.53
17	38.4	1.53
18	3.6	0.71
19	49.1	1.54
20	24.1	1.5
21	1.2	1.08
22	56.5	1.63
23	43.4	1.63
24	54.7	1.62
25	4.9	1.5
26	33.2	1.54
27	56.2	1.5
28	24.3	1.62
29	22.6	1.66
30	20.3	1.63
31	18.4	1.52
32	23.7	1.36
33	8.1	1.57
34	26	1.52
35	27.8	1.5
36	8.6	1.22
37	7	1.71
38	39.1	1.68
39	10.5	1.82
40	10.9	1.66
41	6.2	1.72
42	13.2	1.51

43	38.2	1.47
44	13.1	1.61
45	25.2	1.65
46	7.3	1.42
47	36.3	1.57
48	42.2	1.62
49	10.7	1.5
50	7.3	1.88
51	47.7	1.59
52	28.6	1.67
53	45.4	1.57
54	38.5	1.63
55	35.7	1.54
56	10.8	1.5
57	15.6	1.49
58	30.2	1.64
59	11.8	1.44
60	120.5	1.69
61	20.8	1.74
62	31.9	1.56
63	59.5	1.58
64	21	1.57
65	31.4	1.54
66	49.8	1.65
67	17.8	1.51
68	25.9	1.51
69	20.8	1.56
70	15.3	1.96
71	120.5	1.69
72	63	1.76
73	32.4	1.62
74	15.7	1.98
75	15.4	1.95
76	10.6	1.61
77	19.6	1.57
78	23.2	1.58
79	21.7	1.51
80	29.8	1.65
81	73.5	1.71
82	52	1.63
83	56.6	1.74
84	40.4	1.5
85	18.5	1.43
86	29.7	1.62
87	32.1	1.49
88	36.3	1.53
89	30.3	1.59
90	49.1	1.65
91	13.6	1.31
92	23.9	1.57
93	35.1	1.61
94	43.1	1.64
95	16.8	1.53
96	56.4	1.58

97	6.9	1.53
98	10.8	1.4
99	51.5	1.58
100	35.7	1.54
101	34.8	1.52

Appendix II

IRF-1 Primers position on the transcripts. Different exons are sequentially marked in alternating blue and pink.

Gene: *IRF-1*

Accession number: NC_000005

Transcript length: 11915 bp

Forward Primer

Reverse Primer

ORIGIN

```
   1 TGGAGGGAGA GCTGGGGTGA AGGAAATGAC ACGCCTGGGA GAGTAACTTA CTTCTGCAGG
  61 AGCTTTAGGG AGATGAAGGA AGAAGCCTCC TGGGCCAGAG TTTTGGATGG AAAATGAACA
 121 CCCAGTCAAG TCTCTAGGAC TATACGTGGG GCGGGGACTA GTTGTGCGCG AGAGTTAAGT
 181 AGGGGCCTTA CCAAGGAGCA TGGGACCTGG GCTCCCCAAC CCTTTGGCTA GCCCCATGGC
 241 GTTGATCAGC CCTGAGCTAA TTCCTCCATG CTGCCCAGAA CCTCTCTGGG CCAAGCCCTG
 301 GGGACTCAGA GATGACAGCA ATGCTTCCAT TGCGGAACTC CCATACGCGG GCCACAGGGA
 361 GGCTCTGGAG GCGGCCTGAG GCAAGAGTGC TAGGAGGGAT CAGAGCTAGC CCACCCCTAC
 421 CCTCACTCAG CCGTCTGGGC TTCTCTGAAC CCCTTCTCCT CCTCTGTTCC CTAAAGCCAG
 481 CCAGGGGGAG TCCCAGGGAG GCAGACCGAA AAGGGGTGGG GTGTCATCCT GGTCACTATT
 541 AGACCCTGCA ACGGCGACCT TGAAAACTAC TCAGCGTCTG TTGCCCGAGT GGAGCATAGT
 601 GCTTTACAAT CTCTTCCCAT CACAGCAAAC CATCAAGGTA GGGCTACTAT TATTTTATGG
 661 TTGAAAAACA GAGGTCCTGC GTCCCTTGGG GGCTGTGCCA GCAGCGGCCA AGTTGGGATT
 721 TCCCCTGGTC CAGCAGCCCC AGACAGCACA CGGGGCAGGG TAGGCTTTCT GCCTTCTTCA
 781 CTTCCCCAGG GCAGGTGAGT GACCTGGAGG GAGGGGGTCA CCCCTAAAAA CAGGGGTAGT
 841 GCTAGGACTG AAACCCTCCC TTCTTGATAT CCCACTGGCA AGCTTGAGGA GCCAGGCTGC
 901 CAGTCGGGAG ATTCGGCCCA GTGTTCCCAC TGGAGAGGGC GGCAAGTGCC CGGGCGATCC
 961 CCTCACCTGC GTTCGGGAGA TATACCCCCG CCCCCGCCCC GCCAGGAGGG TGAAAAGATG
```

1021 GCCCCAGGAG CCAGCCGGCT GGGACAAGGC GGAGTGAGAG GACAGGCTGG GGCCAGGGGC

1081 GCTGGGCTGT CCCGGGCAGC CCTCCTCCGG GCAAGCCGGA GCAGGGGTGG ATTGGGAGCG

1141 CTCGGGGCGG GCCCGCGGTG GCCCCGGGGC GGTGGCGCCC GGCCGGAGAG GGTGGGGCGG

1201 AGCAGCCGCC CTGTACTTCC CCTTCGCCGC TAGCTCTACA ACAGCCTGAT TTCCCCGAAA

1261 TGACGGCACG CAGCCGGCCA ATGGGCGCCC GCGCGGCTGT CCGGGGGCGG GGCCGGCCAG

1321 GGCTGGGGAA TCCCGCTAAG TGTTTGGATT GCTCGGTGGC GCCGCTGCCC TGGCAGAGCT

1381 CGCCACTCCT TAGTCGAGGC AAGACGTGCG CCCGAGCCCC GCCGAACCGA GGCCACCCGG

1441 AGCCGTGCCC AGTCCACGCC GGCCGTGCCC GGCGGCCTTA AGAACCCGGC AACCTCTGCC

1501 TTCTTCCCTC TTCCACTCGG AGTCGCGCTC CGCGCGCCCT CACTGCAGCC CCTGCGTCGC

1561 CGGGACCCTC GCGCGCGACC GCCGAATCGC TCCTGCAGCA GAGGTGAGTA CGCCTTTGAG

1621 GCGCGGGGCA CCGGCGGCGT CGAATAAAAG GCGCGCGGGG CACCAGGAAG TGGGGGGTCG

1681 AAAGCTCCAG GCTGGAGACT CGCCGGCGCG CGGCGTTGCC CGGGCCTCCG CGCGGGCTCC

1741 GGGGGGCGCC GGAGGAGCTG CGAGCCGCGG GCCGCGGCGC GGGGAGGGCG GGACGCGGCG

1801 TGGACCGCCC ACCCGGACGA GGCTGCCGGC GCCCGGCAGC TTTCGCAGAT CTGCGTGCGC

1861 GCAGCCGCCA GGGGCCTGTA GGTGGCCCGC TATGTTCGTC CCGCGCATCC ACACGCCGTG

1921 CCGGGGACCG AGTGTCAGCC CACGCGTGGG CGCCCAGTGC TCCCGGCTTT CGGCGGTCCC

1981 AGCTCCGCGC CCAGGCGACA GGTTTTGGGC TCCCTGTGCT GGTGGCAAGG GCTGGCTTAC

2041 TGCCCAGGTG GCTGGAGGGA ATCGTGACCT ACGGAGACTG CGGGAAGAGG CGCCACAGGT

2101 GTTCCTTGGG CCACTTCTCC AGAGGAGGGG AAACCGGGCC GGAAGGGTTA GCGTCCTGGT

2161 CTTAGCGTTG TGGGCGCTGT GGCTGTCAGG AAGGCGTAGA ATGGATTCAG GGGGGCGGGA

2221 GGGGGCTGTT CAGGGTGACG GCTAGCCCTT TGCTAGCTAG TGGTTACAAC TCAAGTCAAG

2281 GGAATTTCTT CTTGGCATCA AGCAAAAGAA GTCCCTCCCT TCCCAAAGGA TTTGAATTTT

2341 GAGCGAAAAG TTCTGAAATT AGGGTATCTG TGCATTTTGT CTCTTTTCCT GCATATGAAT

2401 CCTGAAGCCA TCACTTGCAT GCCTGTCTCC TCCAGAGACT GGCTGGGAGG GGCTGAAGGA

2461 AGGGGCAAAA GCATTTTTGC CTAAGATGCT GAAAAAATTT GGAGAGCAGT TTTATTCCAG

2521 CGCAGCTCCC CTCCGCACTG AGTGTAGTAC CTAGCAGCTG GCTGAGGTGA GGGGAGGGTA

2581 ACTAAGTGAC CTCGGGTGGG GCAGGTCACT GCCCAGGTAC TGTTCAACAG ATTCCAGACT

2641 GGAGCCTCTG TGTTCTCTTT ACAGCCAACA TGCCCATCAC TCGGATGCGC ATGAGACCCT

2701 GGCTAGAGAT GCAGATTAAT TCCAACCAAA TCCCGGGGCT CATCTGGATT AATAAAGTGA

2761 GTGTAACTCT TTGGGTTTTC CTGCCACTGT TTTAACCCAT GTACTTCTGG AGGGACCAAA

2821 GCTTCAGATG CAGCTCAAAA AGGGAAGTGA TAACGGGACA AGCAGGTGTT TCTCCCAGTG

2881 GGTCCTGCAT GCAGGGAGTG TGCACGGCCC AGCCTGGGCC TCACTTGCAT GACTCCTGCC

2941 TTCTTCCCTT CTTGAGGTAG GGCACCCACC TGAAGGCACT TCCAGTTTCC AGCAGCAAGA

3001 CTTTCCAGCA TCTGCAGAGC TGGAGTTCTG CTCTCCTCTA AGCGAGACCC TTACAAACAT

3061 ACACAGCACT CTGCAGGGCT CCAATCGAAC AAATAGAAGA CTGAGAAGTG GATGCTGCTG

3121 GGCAGAAACG TGCCTGGCTT AGCAGAGGAC AAACGAGTTA ATCTTGCACC AGTCACTCTG

3181 GCCCAAGAAG CCTATAGCTG GTGCACTTGG GGCAACATAG ACCCTATAGA CTTAGTAGCA

3241 ATGATAGTAT TCATAATAAT AGCTAATGCT TACTGAACAC TCCCTGTGTG CCTGGCACCT

3301 GCTAAGTATG TTATTTACAT TGTGTCATTT AATCCTCGCA GTAGTCCTGT GGGTTAGATC

3361 TTACTAATGT CATCATTTTC AGATAAGTAA ACAGAGGCAC TGAGAGGTAG ATCATAAGAT

3421 CACACAAAAA GTGATGAAGC CAAGATTTGA ACTTGAACGG TCTGACTCAG AAATCTTTAC

3481 TGTTAACCAT AAGTGATATA ATAACAGTAA GACCTTAGAC TTCATATTTG TCACTGTGTC

3541 CCTACACATC CTCTGGTTTT TAATCCTCAA AATTTTGTTG GATATGTTTT CTCATTTCCG

3601 AGAAGAGAAA ACTGAGGGGC AAAGAGATAC AGTGACAATG CCAGGGTTAC ACAGTGTTCA

3661 CCATCCAAGT CTAGCCCAGA GCTCCCTCAG TGGTATGACC AGGACCCCCT GTGTAAGAGC

3721 CCATGCTCCC AGGTGTCCTG AGGAGTCCTT TCTAATGGAA GAAGTTCTTA CTTCCATGTG

3781 GGTGCTTACA AGCCAGAGAG AAACATCCCA GAGCTTCAAA ACCAGGGCTT TGGGGGAGGG

3841 TGCCCTGTGT GGGTCCTAGC ACATGTGTAA CAGGCAGAGG GAGGTCTTTG TGAGCTAATA

3901 ATGCTGCAGC TCATCCAAAC TAGGTGTCCC TCCTGAGAGA TCCAGAGTGG TCTGTTTAAG

3961 CCAGCCTCAA GATGGGTGTC CAAGCCAGAT GTCAGGGGAA AAAAGGGGAA GTCAGCCTTT

4021 TCTCAGACCT GTCTGGCTGG GCAGGCCTGG GTCTCAGACT CAGCCCCAAA GTCTGTGGTC

4081 TCTGACCTGA CACAGCCTTA TGTGTATGTG TGTATTGTTC AGGAGGAGAT GATCTTCCAG

4141 ATCCCATGGA AGCATGCTGC CAAGCATGGC TGGGACATCA ACAAGGATGC CTGTTTGTTC

4201 CGGAGCTGGG CCATTCACAC AGGTGTGTGC CTGGGACTCA GGCCTAGGAA GCCCAGGGTA

4261 GAGACAAGAG GAGGCACTCA CGTTAACACA GAGGCTCTTC ACTGGGGTCC CTGAGCTCCC

4321 TGAGACAACA TGCAGAATTA CTGGGAAGAG GGGCTGGTGG CAGACTTGTG TTTCTGGAGA

4381 AGAGAGTCGA TCATCTCAGC AAATTCTCAA AGGGAAAAGC CAAGATCTTA GAAAGTGTGT

4441 GGCTTCAGGG GGTTTGTGGC TAGATGAAAG TTCTCCCTGG CAAAAGCATC TGTGAAAAGC

4501 AGCTGTAAGC CAGGGCACTG AAAGAGACCC AGGTCTGCCT TTTTCTTCGT GTTGACCAAG

4561 GCCCTTGGTC CAAGCCTCAT GTGGTTGGTG GCCTCCTTTA TCCTTGAGAG ATGGAGCTCT

4621 AGGCCCATCT CAGAACAGTC AGCCCACCCA TTTAGTAACT GTTCTCTGCT GCCCAGTCTG

4681 TGCCCACTCT ACCCTCTGGC TGCTGATAGC CCAAGGAGGA AGACTGGGCA TAGTCTGAGA

4741 CACAGATAGT ACACTTTGGG GATATGGGGA CTCTAGTGCT TCTGGCTGGG CCCTTCACTG

4801 AGGCCCGCTA GATGTGTTTA AGCCAAGCCT GGGCATTTGA GAAGGCCCAG GGCCTAGGAC

4861 CTGCAGAGTG TCACCGGGAG TACCTGCTGG TTTGACCACT GTGGCTCTCT GGTAGCATAA

4921 GAGGTCAGGG GTACCTTGCC TTCCTCCTTC AGGCCAGGGG CAGCTGAGGA TCCCTACCCA

4981 TGGCCCTGAC GATCCTCTTT TTTCTCCTGC CCTCTAGGCC GATACAAAGC AGGGGAAAAG

5041 GAGCCAGATC CCAAGACGTG GAAGGCCAAC TTTCGCTGTG CCATGAACTC CCTGCCAGAT

5101 ATCGAGGAGG TGAAAGACCA GAGCAGGAAC AAGGGCAGCT CAGCTGTGCG AGTGTACCGG

5161 ATGCTTCCAC CTCTCACCAA GAACCAGAGA AAAGGTATCC AAGGACTCTG GGTCCTTGGG

5221 AAGCCCTCAG GGAGGGAGGG TAGAAGGAGG TCAGCTGGGG CTGGAGAGCC TGCACCAAGG

5281 CTGACAGCCC GTCTGCCCCA CAGAAAGAAA GTCGAAGTCC AGCCGAGATG CTAAGAGCAA

5341 GGCCAAGAGG AAGGTGAGTG TGGTCCTAAG CAGCCAGGCC TTTGGTCACC TGTGGGCCAG

5401 GGTGAGCAGT GGAAGAAATG CTAAGGTGGG CCTGGGCCTA AGCTGCTTTC TCCCTCGACA

5461 GTCATGTGGG GATTCCAGCC CTGATACCTT CTCTGATGGA CTCAGCAGCT CCACTCTGCC

5521 TGATGACCAC AGCAGCTACA CAGTTCCAGG CTACATGCAG GACTTGGAGG TGGAGCAGGC

5581 CCTGACTCCA GGTGAGCTGG TCCAGGTCTG GCAGGAGACC CCACAGGTCA GTGGGATGAC

5641 TCTTTCTCTT GGAGGCATGG TGCTGGCACA TGGTGGCCCA TTAGTGCAGG CTGCAGGGTT

5701 GGTCGGAGGG CGCTCGATGT CTTGCAAACT AAGAAAGCAC ACAACCTTGA CCTGTGGCTT

5761 CTGCTGTTCC CCAGCACTGT CGCCATGTGC TGTCAGCAGC ACTCTCCCCG ACTGGCACAT

5821 CCCAGTGGAA GTTGTGCCGG ACAGCACCAG TGATCTGTAC AACTTCCAGG TGTCACCCAT

5881 GCCCTCCACC TCTGAAGGTT GGTGCTCCTG GGGCCTGGCC TGCCTGCTTG ACTGTCTGGG

5941 TCCTGTGAAG GGCTTCCTGA GAGAGAAAAG ATGATCAGAA CTCCACCTGG CACTGAATTG

6001 ATTGAGTTGG GCATTGCCCA GTCTTAGCCA CCATAGGGGG AGGCAAGCGA CGGGGACACT

6061 AGGAAGGCAG TTCAGAGTGG GCTGCAGTAC AGTGGGGGCT GGTGAGAGGA GGGAAGGGGG

6121 CCAGGGGCTG CATTTTGGGG TGCTGGTTCT CCTTCCTCCT CTGTAGCCCA GCATCTGAGG

6181 GTGAGGAAGG AAGTAGGGTA GGGGTGGGAA GCGGCGTGGC TTCAGGGTTT GAGAGGCTGA

6241 GTCACCAGGC CAGGGTCCTG TTCTGGAATC TCTATGGCAG ATAGGTCCAC CGGGAGGGTG

6301 TGTGTGTGTG TGTGTGTGTC AGAGAGACAG AGAGACAGAG AAAGGGCAGG GGGATCTGGT

6361 GGGCTGGAAC TGGAACTGCA GGGTGAGTGT GGCTGACTGC CAGCCAACCT CTCTGCTTTC

6421 CCCATCCACA CGTACAACAG ATGAGGATGA GGAAGGGAAA TTACCTGAGG ACATCATGAA

6481 GGTAAAGCCC CTTCCTACCT GGGCACTCTT GAAGTGACCG TTTCTCAGTG AGGAGAGAGA

6541 ACCAGTGAAG CTTCCAAATC AGAGGATGGG TAGCTGCTGT TGTCACCTGG CTGCTTGCAT

6601 TGTCCCACAA GTGCCACATT CACGTGGCTT GACTGGTGGG AAAGCCACCA TGGGAAGGGA

6661 AGGCAGGTGG GAGGCCTGGC CTCTGACAGG CCGTCCTGAA GCAAGCCTTG GGGCATCAGA

6721 CAGCTCTGTG AGTCAGGCAC TATCAGCGAT GGGTCCCTGG CCTGCATCCT CTGCCCCAAC

6781 ATGCCCCAGC CCTGCTAGTT CGGGAAATGC ACATCAGGCT TCAATAATCA GCCTTTAGGA

6841 TCCGTTAATA TGATGATGGC TTTATAGAAA AAGTTAGCAA ATTATCCTCC AGGTTTTTTT

6901 TTCTGCTTCA GTTTTGAAAG TGAATATAGT TTTTGCAGCC GGGGGCAGTG GCTCATGCCT

6961 GTAATCCCAG CACTTTGGAA GGCGAAGGTG GGTGGATCAC CTGAGGTCAG GAGTTTGAGA

7021 CCAGCCTGAC TAACATGGTG AAACCCATCT CTACCAAAAA TATAAAAATT AGCTGGGCCT

7081 GGTGGCGCAT GCCTGTAATC CCAGCTACTC TGAAGGCTGA GGCAGGAGAA TCGCTTGAAC

7141 CTGAGAGGCG GAGGTTGCAG TGAGCTGAGA TTGTGTCATT GCACTCCAGC CTGGGCAACA

7201 AGAGCAAAAC TCCATTTCAA AAAAAAGTTT TTGCAGTAGT TGTACGCCAG CTGTTCCATT

7261 AGCCCAAAAA ATTGAGACAT GGATGTCGTT CCTTATCTCT AGCTTTTCTA GTCATCTTTT

7321 CTTGATTTAT TATGCTAACC TTTGTTTTAA GCCACATTCC CTCTTACTAT GTCCTTACAC

7381 AGTTGAGAGG GAAGTCGTGG AGATGCTATA CCAGAGAGTG GGTGTGAGAG GGGTGGGAAA

7441 ATGAATTGAG GACCAGTGCC AACATGCATT TCTGCCTCCT CCCTCCCGGG CCCTTGTCCT

7501 GACTGCAGTG CACTTCTGCA TCCTATCTGA GATTGTGAAA ATGGCCAAGG GTGTGATACT

7561 GGCTGAGAGG AGCTGGCTCA TTGAGGGCAG GGCCACAGGG TGAGTCTGCA CTGGAAGGGA

7621 GTTGATAGCC TCTTGCTCTT CTGTCCCCAG CTCTTGGAGC AGTCGGAGTG GCAGCCAACA

7681 AACGTGGATG GGAAGGGGTA CCTACTCAAT GAACCTGGAG TCCAGCCCAC CTCTGTCTAT

7741 GGAGACTTTA GCTGTAAGGA GGAGCCAGAA ATTGACAGCC CAGGGGGTAA GAAGGCCCTG

7801 GATCCTTATG GCTTCTTAGA TGAGGGAGAA CCACGTAGGG ATGGAGAAAG CTTGGGGGCA

7861 GGGCCAGGGA GCAGGGCGGT AAAGCATCTG GGGTACTGAC ACATTGTGAA TTAGCTACGG

7921 CTGCCATGCC TTAAGGTTTG CCTGAAGCTG AGTGGATGTT TACTGCTGTG CTGGGAAGAG

7981 CAGAGGCCAT GTCTATGGCC TTCAGGGGTA GGGGGAAGCA CACCTGATGC CACCGTCCCC

8041 TACCCTCATA CAACCTTCTT CACATCTTCT AGGGGATATT GGGCTGAGTC TACAGCGTGT

8101 CTTCACAGAT CTGAAGAACA TGGATGCCAC CTGGCTGGAC AGCCTGCTGA CCCCAGTCCG

8161 GTTGCCCTCC ATCCAGGCCA TTCCCTGTGC ACCGTAGCAG GGCCCCTGGG CCCCTCTTAT

8221 TCCTCTAGGC AAGCAGGACC TGGCATCATG GTGGATATGG TGCAGAGAAG CTGGACTTCT

8281 GTGGGCCCCT CAACAGCCAA GTGTGACCCC ACTGCCAAGT GGGGATGGGG CCTCCCTCCT

8341 TGGGTCATTG ACCTCTCAGG GCCTGGCAGG CCAGTGTCTG GGTTTTTCTT GTGGTGTAAA

8401 GCTGGCCCTG CCTCCTGGGA AGATGAGGTT CTGAGACCAG TGTATCAGGT CAGGGACTTG

8461 GACAGGAGTC AGTGTCTGGC TTTTTCCTCT GAGCCCAGCT GCCTGGAGAG GGTCTCGCTG

8521 TCACTGGCTG GCTCCTAGGG GAACAGACCA GTGACCCCAG AAAAGCATAA CACCAATCCC

8581 AGGGCTGGCT CTGCACTAAG AGAAAATTGC ACTAAATGAA TCTCGTTCCC AAAGAACTAC

8641 CCCCTTTTCA GCTGAGCCCT GGGGACTGTT CCAAAGCCAG TGAAATGTGA AGGAAAGTGG

8701 GGTCCTTCGG GGCGATGCTC CCTCAGCCTC AGAGGAGCTC TACCCTGCTC CCTGCTTTGG

8761 CTGAGGGGCT TGGGAAAAAA ACTTGGCACT TTTTCGTGTG GATCTTGCCA CATTTCTGAT

8821 CAGAGGTGTA CACTAACATT TCCCCCGAGC TCTTGGCCTT TGCATTTATT TATACAGTGC

8881 CTTGCTCGGC GCCCACCACC CCCTCAAGCC CCAGCAGCCC TCAACAGGCC CAGGGAGGGA

8941 AGTGTGAGCG CCTTGGTATG ACTTAAAATT GGAAATGTCA TCTAACCATT AAGTCATGTG

9001 TGAACACATA AGGACGTGTG TAAATATGTA CATTTGTCTT TTTATAAAAA GTAAATTGTT

9061 TATAAGGGGT GTGGCCTTTT TAGAGAGAAA TTTAACTTGT AGATGATTTT ACTTTTTATG

9121 GAAACACTGA TGGACTTATT ATTGGCATCC CGCCTGAACT TGACTTTGGG GTGAACAGGG

9181 ACATGCATCT ATTATAAAAT CCTTTCGGCC AGGCGCGGTG GCTCACACCT GTAATCCCAG

9241 CACTTTGGGA GGCCGAGATG GGTGGATCAC CTGAGGTCAG GAGTTCGAGA CCAGCCTGGT

9301 GAAACTCCAT TTCTACTAAA AATGCAAAAA TTAGCTGGGC GTGGTTGCGG GTGCTTGTAA

9361 TCCCAGCTAC TCAGGAGGCT GAGGCAAGAG AATCGCTTGA ACCTGGGAGG TGGAGGTTGC

9421 AGTGAGCCGA GAACATGCCA TTGCACTCCA GCCCGGGCAC CAAAAAAAAA AAAAAAAAAA

9481 AAAACCTTTC ATTTGGCCGG GCATGGTGGC TTATGCCTGT AATCCTGGCA CTTTGGGAGG

9541 CCAAGGTGGG CAGATCACCT GAGGTCAGGA GTTTGAGACC AGCCTGGCCA ACATGGTGAA

9601 ACCTCATCTC TACTAAAAAT ACAAAAATTA GGCCGGGCAC GGTGGCTCAC GCCTGTAATC

9661 CCAGCACTTT GGGAGGCAGA GGCGGGCGGA TCACGAGGTC AGGAGATCAA GACCATCCTG

9721 GCTAACACGG TGAAACCCCG TCTCTACTAA AAATATAAAA AATTAGCCGG GCCTAGTGGC

9781 GGGTGCCTGT AGTCCCAGCT ACTCGGGAGG CTGAGGCAGG AGAATGGCAT GAACCCCGGA

9841 GGCAGAGCTT GCAGTGAGCC GAGATTGCAC CACTGCACTA CAGCCTGGGC GACAGAGCGA

9901 GACTCCGTCT CAAAAAAAAA AAAAAAAATT AGCCGGGCCT GGTGGCGGGC GCCTGTAATC

9961 CCAGCTACTG TGGAGGCTGA AGCACAAGAA TCACTTGAAC CCGGGAGATG GAGGTTGCAG

10021 TGAGCTGAGA CTGTGCCACT GCACTCCAGC CTGGGTGACA AGAGTGAGAC TTTGTCTCAA

10081 AAAAAAAAAA ATCCTTTTGT TTATGTTCAC ATAGACAATG GCAGAAGGAG GGGACATTCC

10141 TGTCATAGGA ACATGCTTAT ATAAACATAG TCACCTGTCC TTGACTATCA CCAGGGCTGT

10201 CAGTTGATTC TGGGCTCCTG GGGCCCAAGG AGTGTTAAGT TTTGAGGCAT GTGCCATAGG

10261 TGATGTGTCC TGCTAACACA CAGATGCTGC TCCAAAAAGT CAGTTGATAT GACACAGTCA

10321 CAGACAGAAC AGTCAGCAGC CCAAGAAAGG TCCTCACGGC TGCTGTGCTG GGTAGCACTT

10381 GCCATCCAGT TTCTAGAGTG ATGAAATGCT CTGTCTGTAC CGTTCAATAC AGTAGGCACT

10441 GGCACTAGCC ACATGTGCCA GCTAAGCACT TGAAATGTGG CCAGTGCAAT AAGGAATTGA

10501 ACTTTTAATT GCATTTAATA AACTGTATGT AAATAGTCAC ATGTGGTCAG TGGTTACCAT

10561 ATTGAACAGT GCAGGTAGAT ACTGGACTGG GGGCAGATCT GAGGGAGAGG GGTTTGAGTA

10621 GTGGGAGGAC ACTGGGGATA GGGGCTTGGG GCTATTTACC TGCCATTTTG AGTAGTTTGC

10681 TATTTTAGCA GCCAACAATA ACTATTGGTG CTGAATACCA GCCCTGCAGT GTAGCATGAG

10741 ACAGGTCCAT GCACACATGC ATTAGGAAAA CACCTTCATG AAGCAGGATT CTGCCTGGGC

10801 TGATGCACAC AACCTCTATG GAGGGTAAAC AGTGTTTCTG AAGACCGTAG TTTGGGAACC

10861 CCTGACATAT GACAATGCCC CCTTAGATAA GCTCAAGTTA CAGGAATGTC TGAGGGTGGA

10921 AGGTGTGGAT ATGTGCTTTT CCTGTCTCCC TCTTCAGTGT CTGGCCATGG GGCATAAACA

10981 CTACCCAGCA GTAGGTAGGC TGGCCAAGAG AAGCCAGCTT GCATCACCAG CATCATCTAG

11041 GGAATGGAAT CATGGCAGTA ATACGTTGCT TAGGAAACAA AAGCTCTATG GACACATCTT

11101 CCACCTTCTC AGTCCCAGAA ACCATATGTA CTGTGACCCC GCTCACTAGG CCCAGCCCTC

11161 GGGAAGAGTG TGGGCCCTTG AAAAGGGAAG ACTGAGTGAG AAAATGATGA GAAAACTACA

11221 AAATGGGCAG AGGTCAGTCT GACACATTCA TTCTCTGTCA AGCTCAGGAA GTACTGGTCC

11281 CTGATCTTGG AGATGCTGTG TGAGTGGCAG GGGGACTCCT GCTGGGTAAA TATTCTATAT

11341 GTGGATGCCT GGACAGGCCC CTATCCCAGG CCCTGCTTGT CAGAAGCTCC CCTTGGGCCG

11401 AGCGCGGTGG CTCACACTTG TAATCTTGGC ACTTTGGGAG GCCGAGGCAG GTGGATTGCC

11461 TGAGTTCAGG AGTTCAAAAC CAGGCTGGGC AACATGGTGA AACCCTGTCT CTACTAAAAA

11521 AAAACTAACC AGGCGTGGTG GTGCATGCCT GTAATTCCAG CTACTAGGGA GGCTGAGGCA

11581 GGCCAATCAC TTGAACCCAG GAGGTGGAGG TTGCAGTGAG CTGAGATCAC GCCACTGCAC

11641 TCTAGCCTGG GCAACAGAGC GAGACTCTGT CTCAAAAAAA AAAAAAAAAA AGAAGTTCTA

11701 CTTGGAAGCT CCACTTGGAT TTCTCAAGAA TAGCTTCACC TGGGAACAGA GGAATAGACA

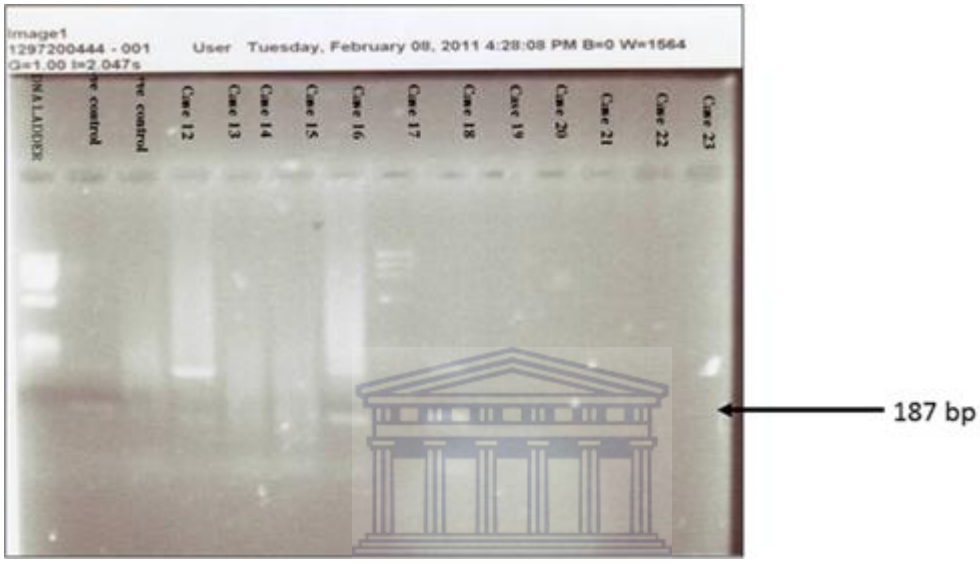
11761 GGATGGACTT TTCCAGCTCC TTCAGGGACC AGCCCTTTTT AAGATTTGGA TTGAGGTGGC

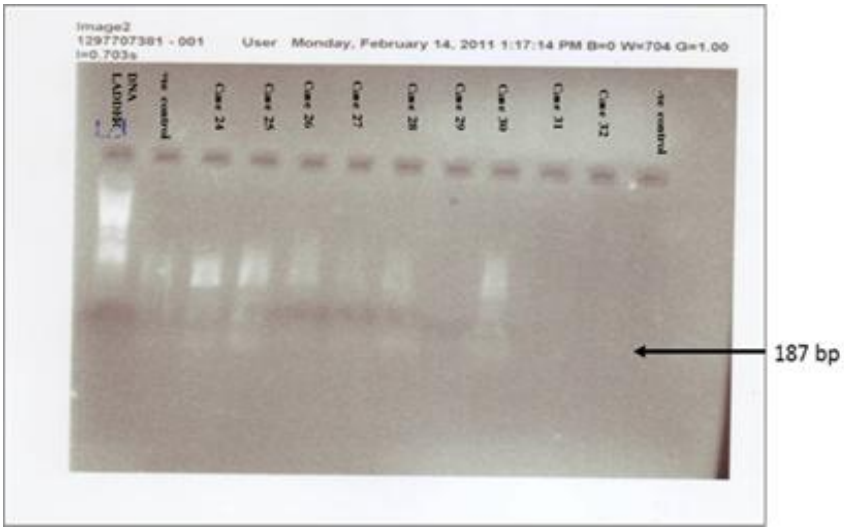
11821 TAGCCACCTG TGGCTTCCAT CTGGGTTCTC CTAGTGGGTG ATGGCAGGTG GTGCAGAGCA

11881 AGGTAGAGTG GACTGACGGG AGGAAAGTGA TACCA

Appendix III

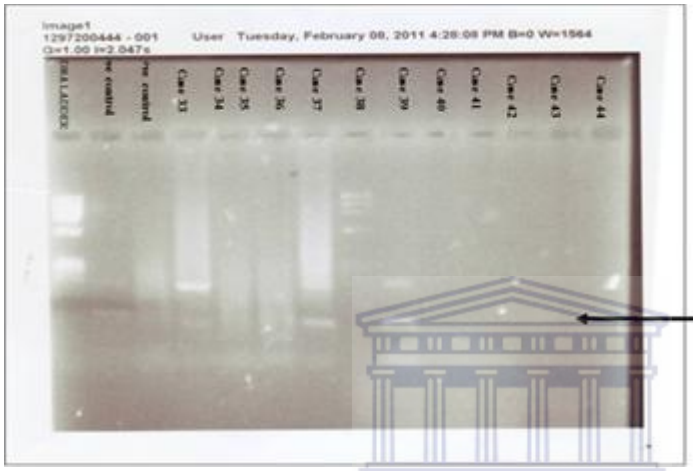
Expression profile of *IRF-1* of Cases 12 to 32 on tumor tissue on lanes 4 to 14, 100bp marker in lane 1 and negative and positive control respectively. PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK). All cases were tested at 35 PCR cycles. The arrow indicates expression of *IRF-1* at 187 bp.

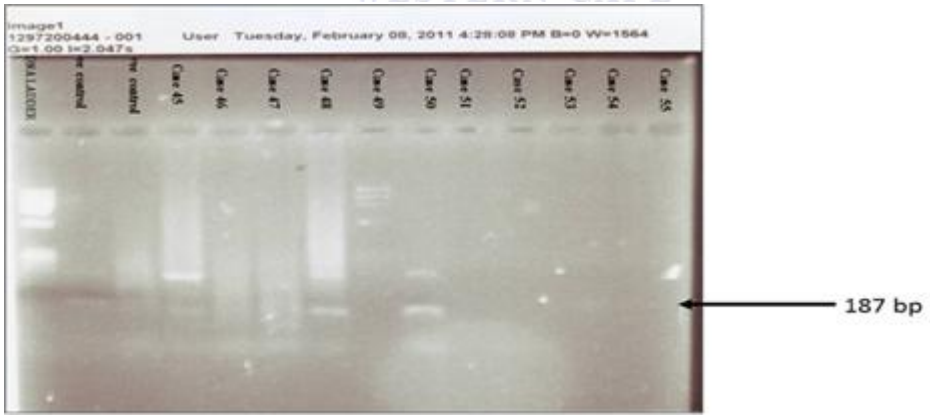




Appendix III

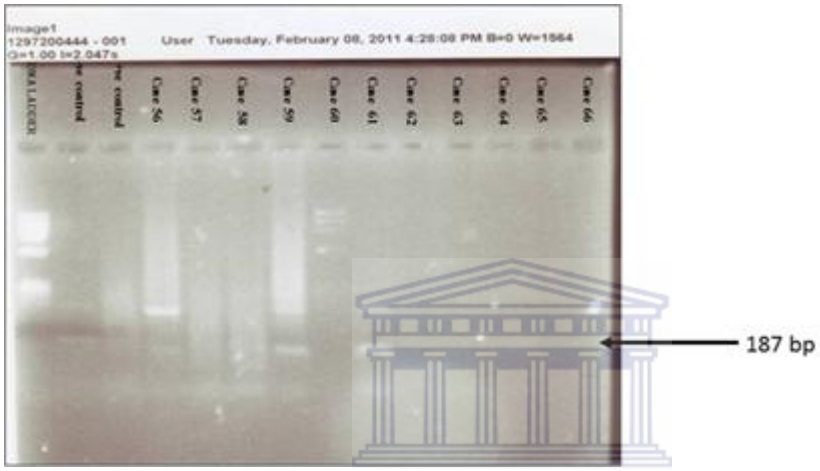
Expression profile of *IRF-1* of Cases 33 to 44 on tumor tissue on lanes 4 to 14, 100bp marker in lane 1 and negative and positive control respectively. PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK). All cases were tested at 35 PCR cycles. The arrow indicates expression of *IRF-1* at 187 bp.

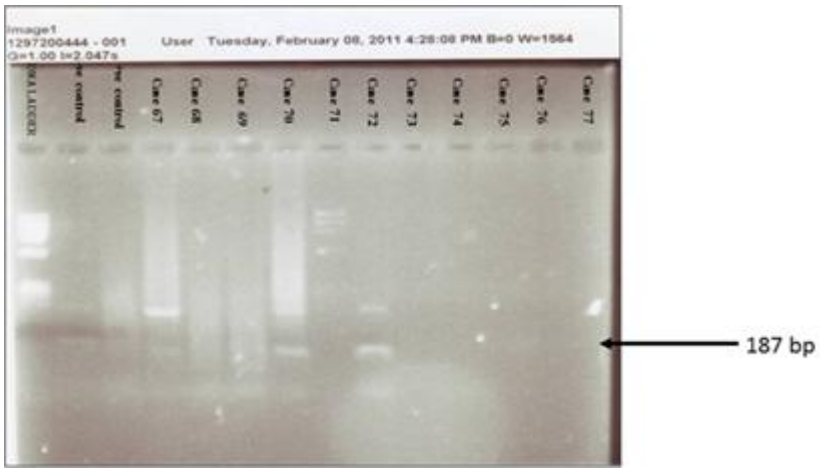




Appendix III

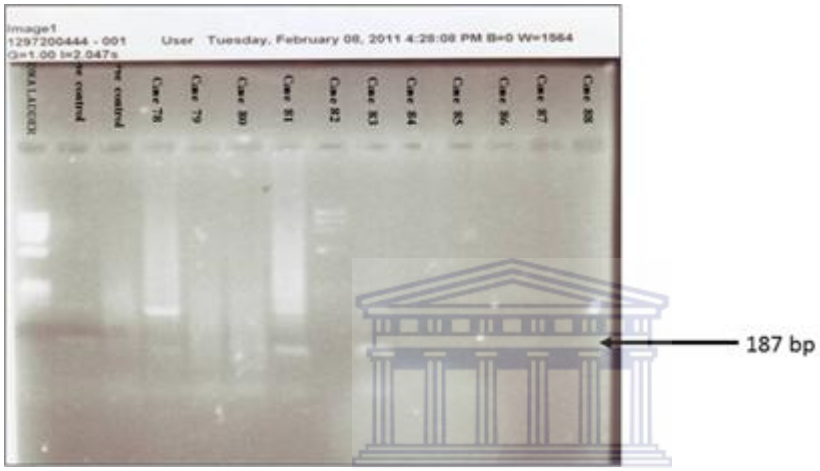
Expression profile of *IRF-1* of Cases 56 to 66 on tumor tissue on lanes 4 to 14, 100bp marker in lane 1 and negative and positive control respectively. PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK). All cases were tested at 35 PCR cycles. The arrow indicates expression of *IRF-1* at 187 bp.

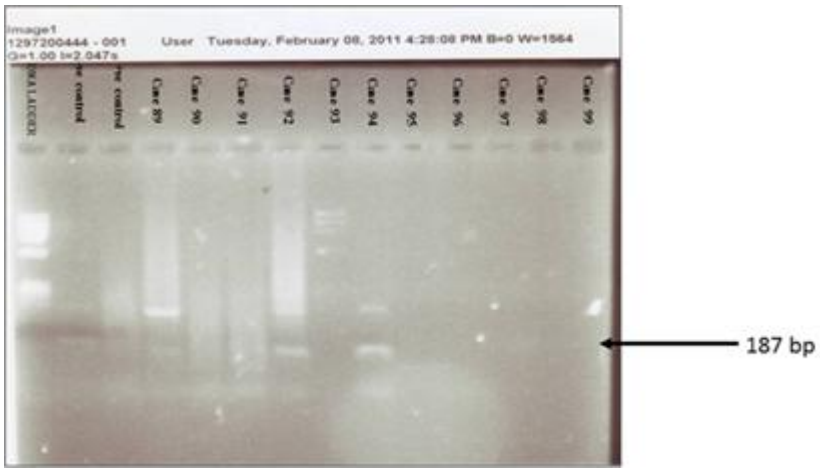




Appendix III

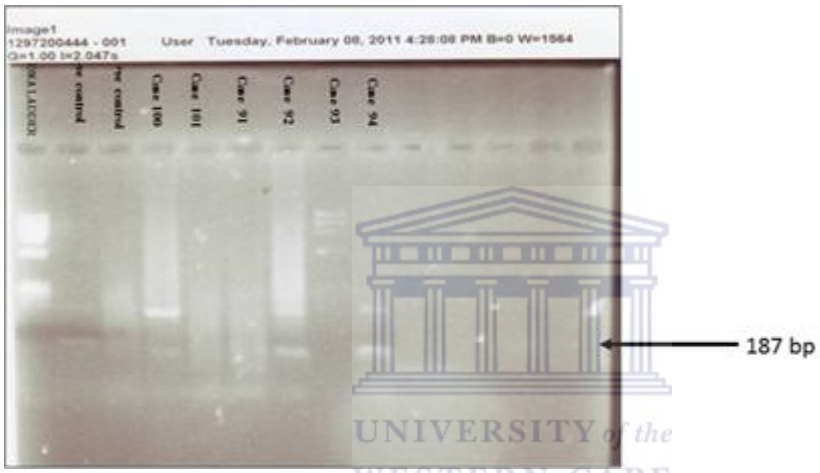
Expression profile of *IRF-1* of Cases 78 to 99 on tumor tissue on lanes 4 to 14, 100bp marker in lane 1 and negative and positive control respectively. PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK). All cases were tested at 35 PCR cycles. The arrow indicates expression of *IRF-1* at 187 bp.





Appendix III

Expression profile of *IRF-1* of Cases 100, 101, 91, 92,93 and 94 on tumor tissue on lanes 4 to 14, 100bp marker in lane 1 and negative and positive control respectively. PCR samples were amplified on a Multi Gene II gradient thermal cycler (Labnet, UK). All cases were tested at 35 PCR cycles. The arrow indicates expression of *IRF-1* at 187 bp.



Appendix IV

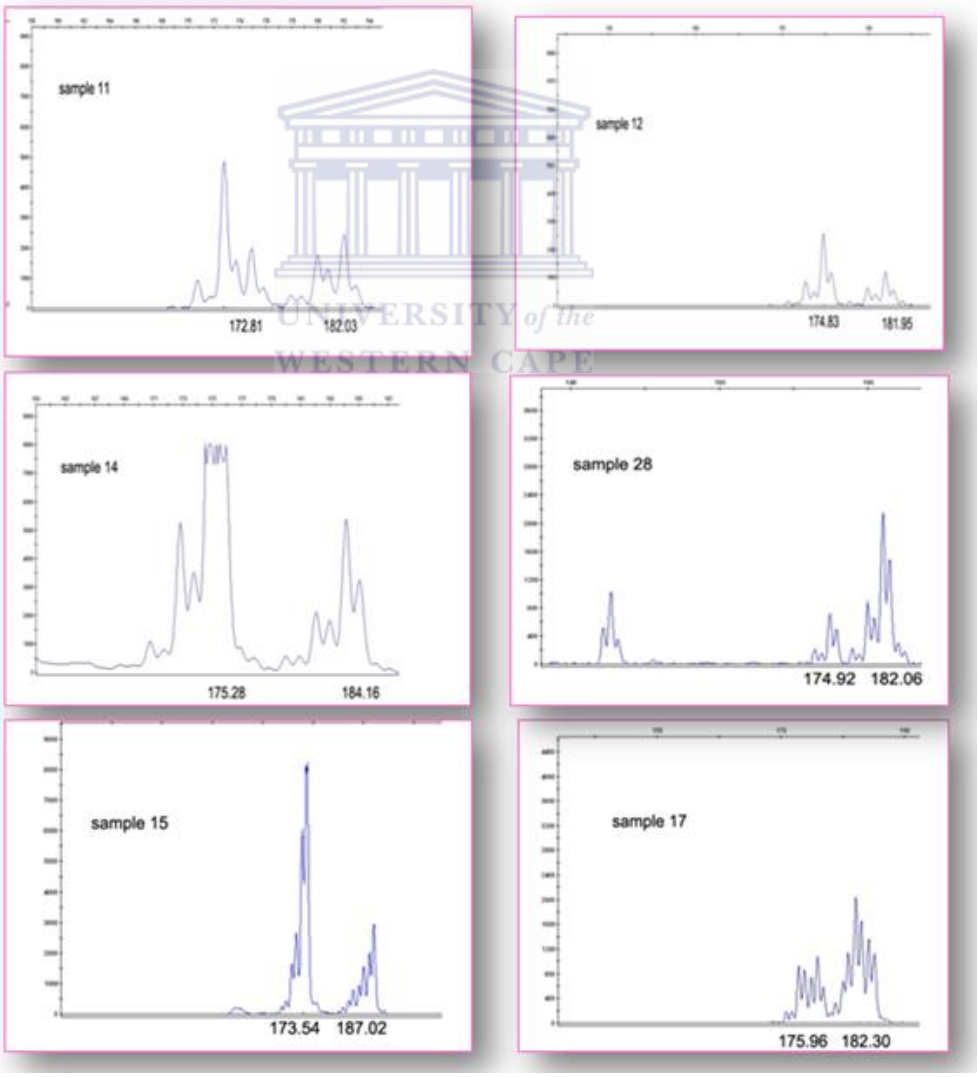
ABI 377 Sequencer (Adapted from **Azco Biotech, Inc. catalogue)**

ABI 377 Sequencer

Appendix V (a)

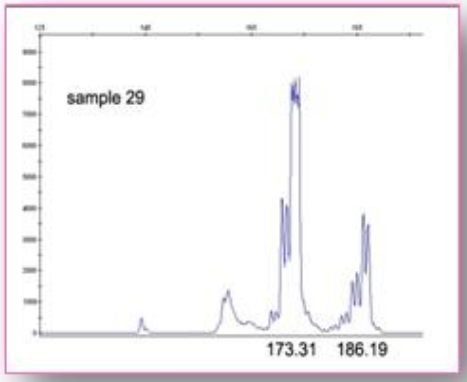
Samples with Loss of heterozygosity:

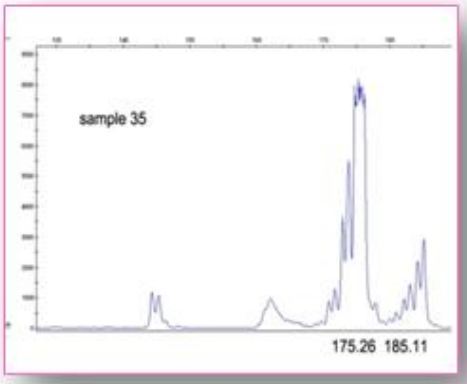
Allele sizes were determined by electrophoresis of PCR products in 6% denaturing polyacrylamide gels and were compared to ROX 500 size standards (Applied Biosystems, Foster City, CA), using an automated sequencer (ABI 377), according to the manufacturer's instructions (Applied Biosystems). Fluorescent signals from the different size alleles were recorded and analyzed using GENOTYPER version 2.1 and GENESCAN version 3.1 software (Applied Biosystems).

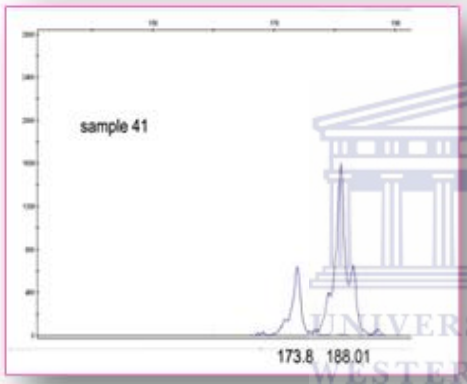


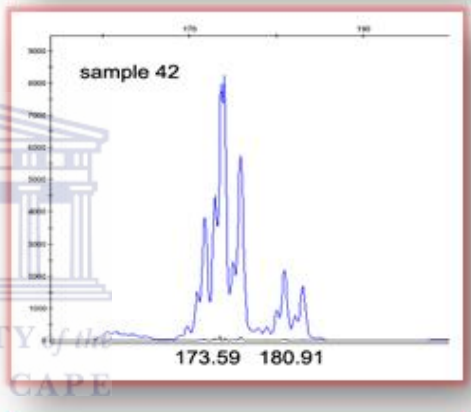
Appendix V (a)

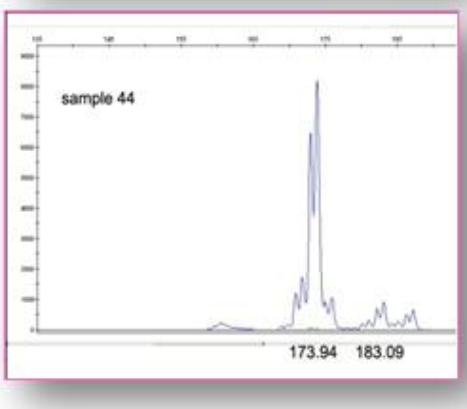
Samples with Loss of heterozygosity continued:

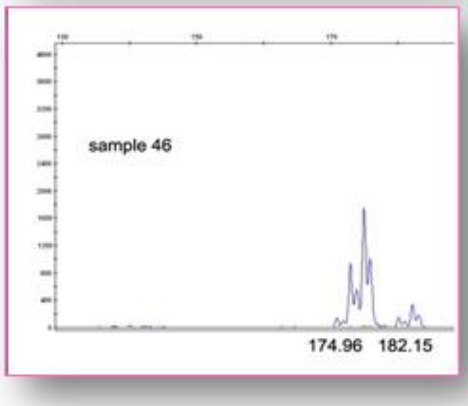






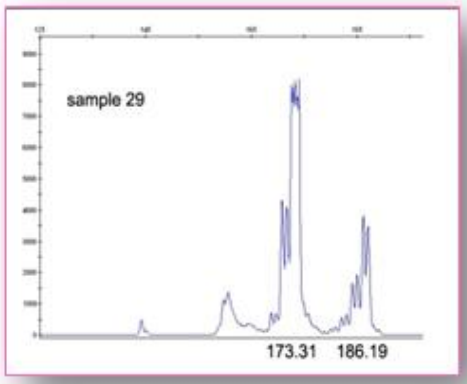


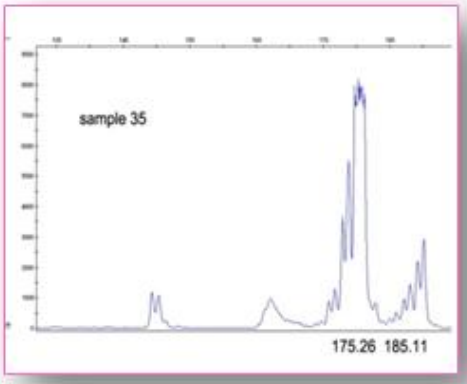


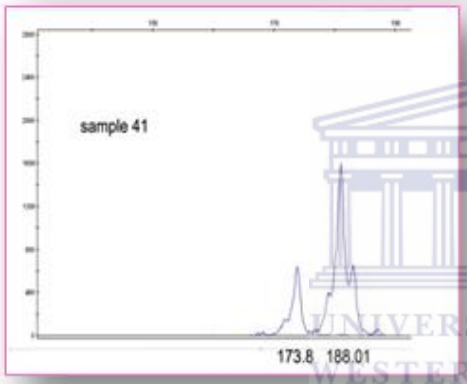


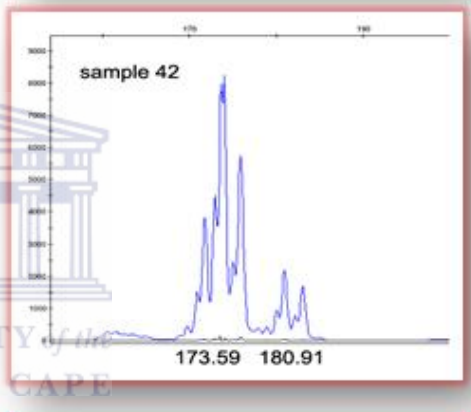
Appendix V (a)

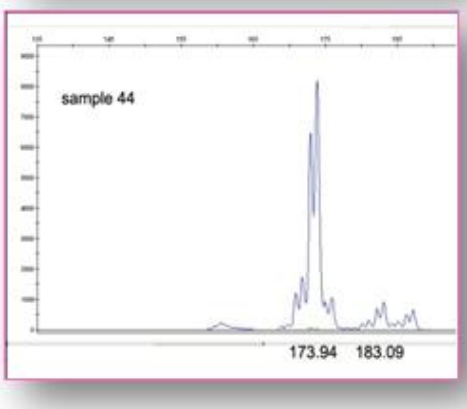
Samples with Loss of heterozygosity continued:

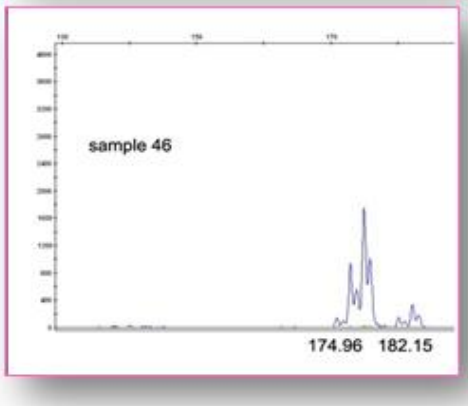








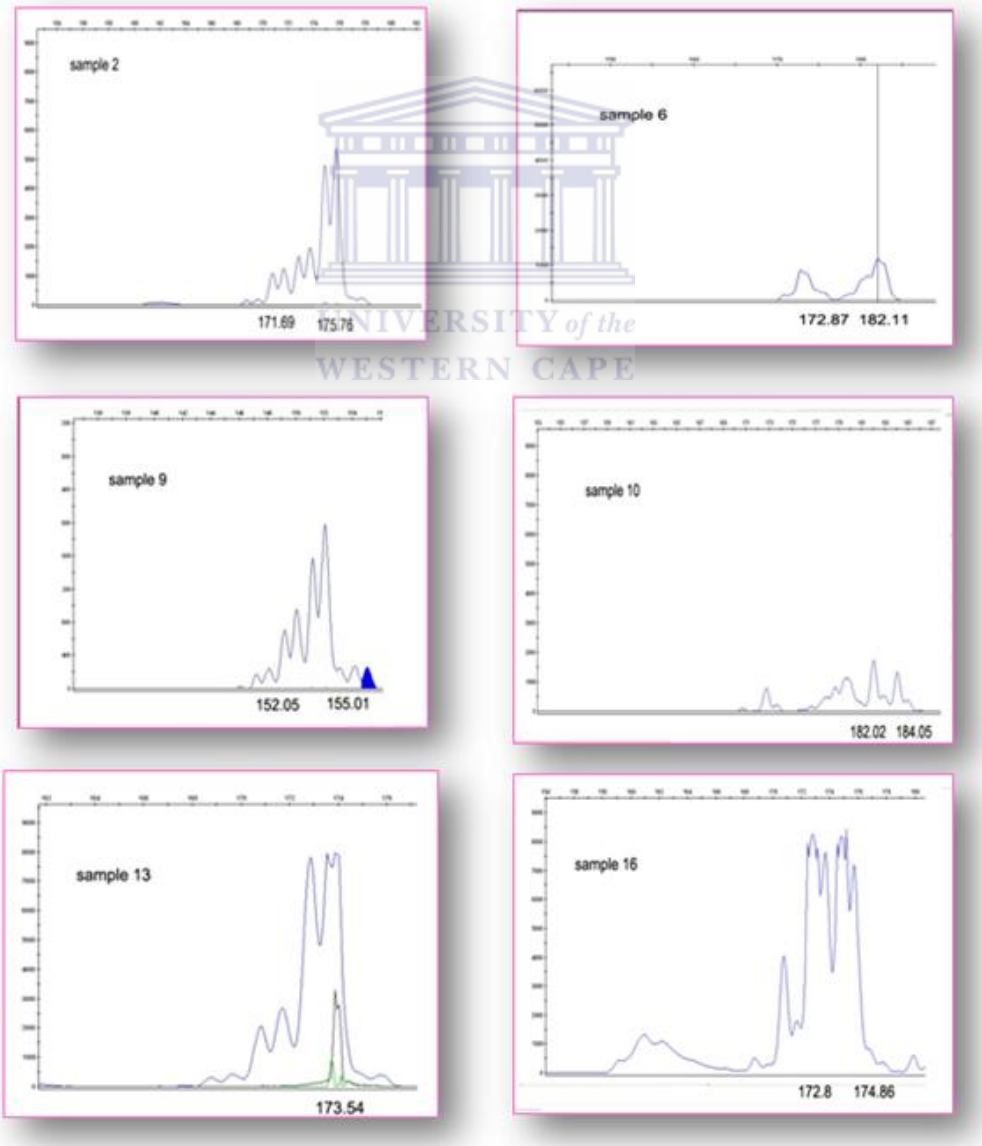




Appendix V (b)

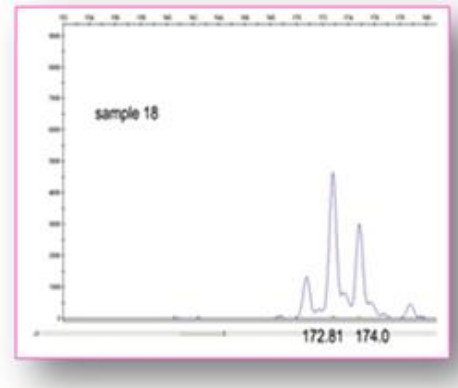
Samples with no loss of heterozygosity:

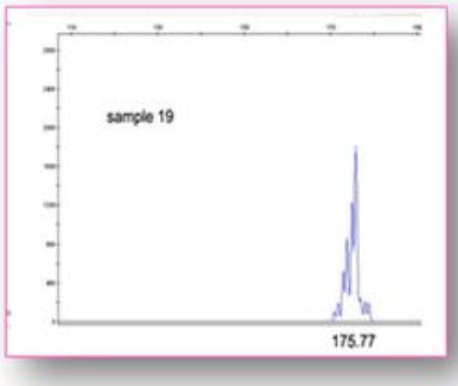
Allele sizes were determined by electrophoresis of PCR products in 6% denaturing polyacrylamide gels and were compared to ROX 500 size standards (Applied Biosystems, Foster City, CA), using an automated sequencer (ABI 377), according to the manufacturer's instructions (Applied Biosystems). Fluorescent signals from the different size alleles were recorded and analyzed using GENOTYPER version 2.1 and GENESCAN version 3.1 software (Applied Biosystems).

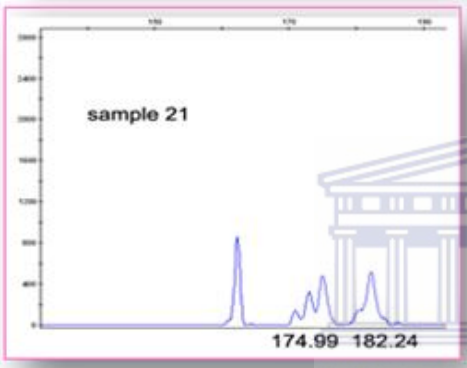


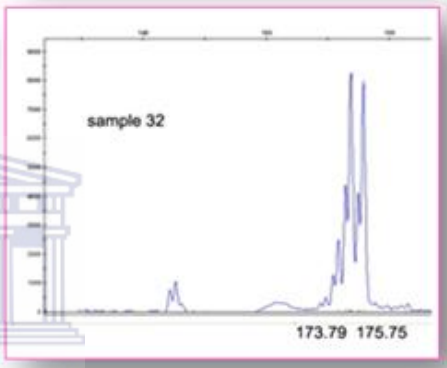
Appendix V (b)

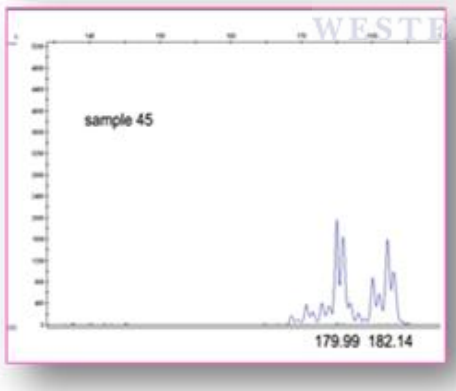
Samples with no loss of heterozygosity continued:

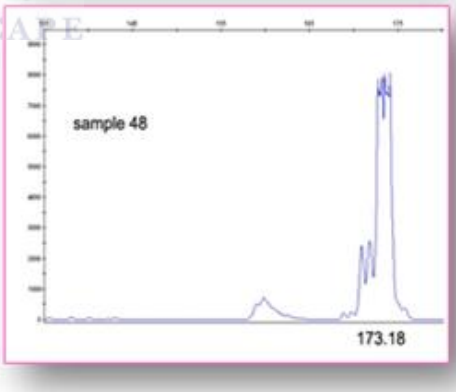








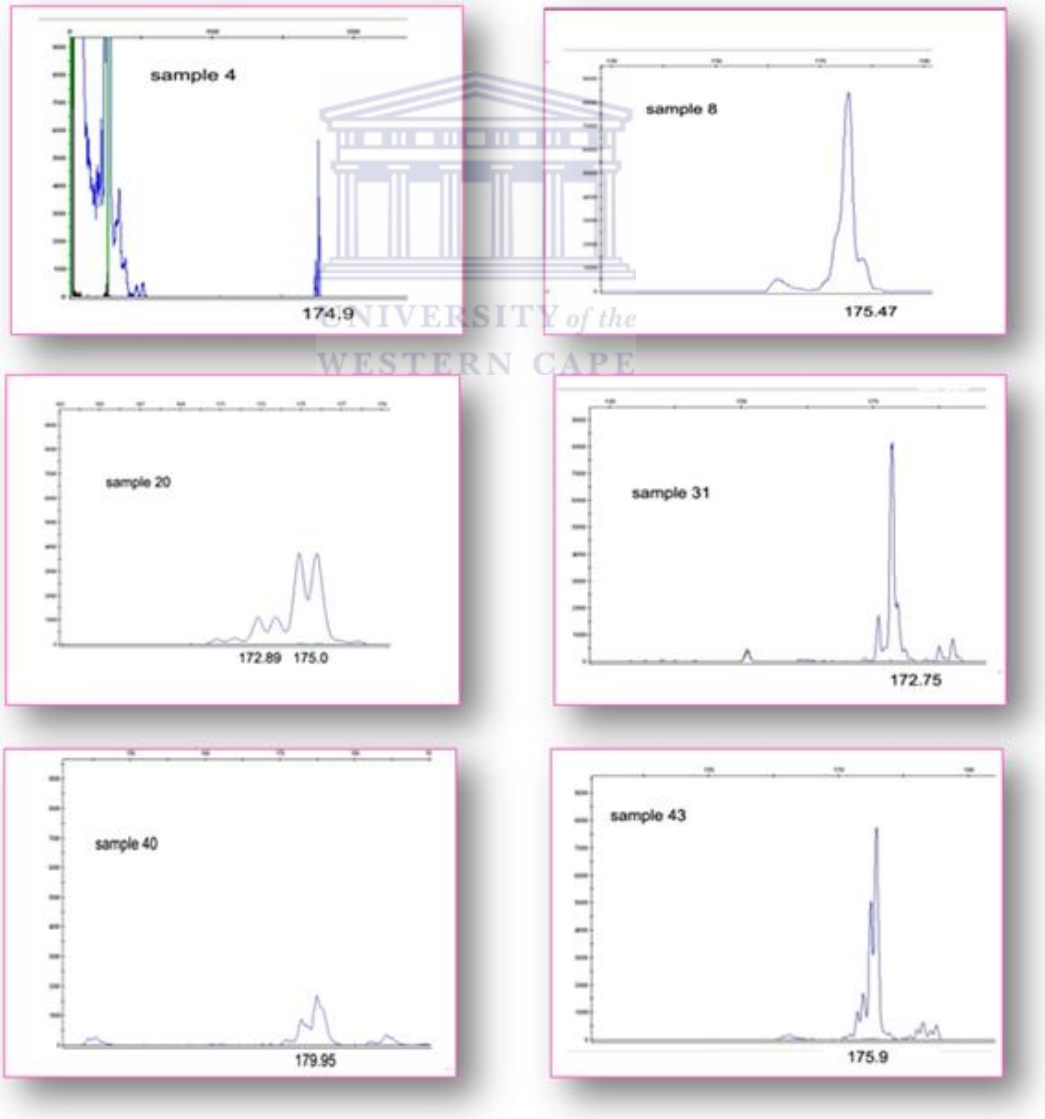




Appendix V (b)

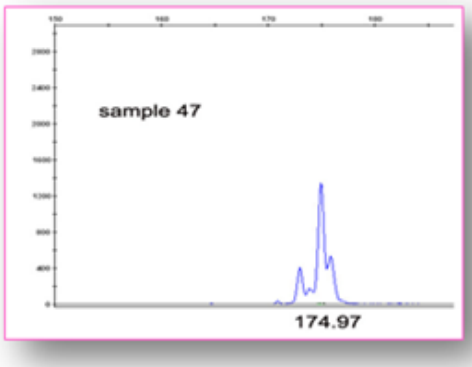
Non Informative samples:

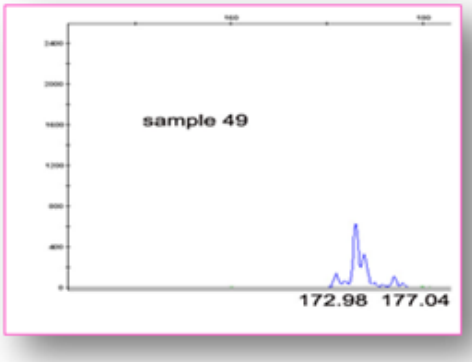
Allele sizes were determined by electrophoresis of PCR products in 6% denaturing polyacrylamide gels and were compared to ROX 500 size standards (Applied Biosystems, Foster City, CA), using an automated sequencer (ABI 377), according to the manufacturer's instructions (Applied Biosystems). Fluorescent signals from the different size alleles were recorded and analyzed using GENOTYPER version 2.1 and GENESCAN version 3.1 software (Applied Biosystems).

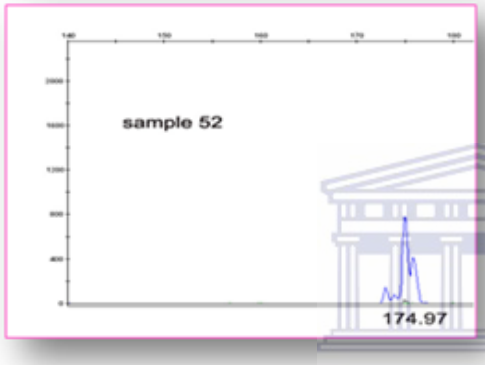


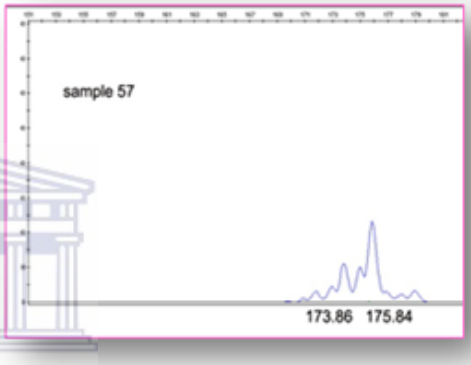
Appendix V (b)

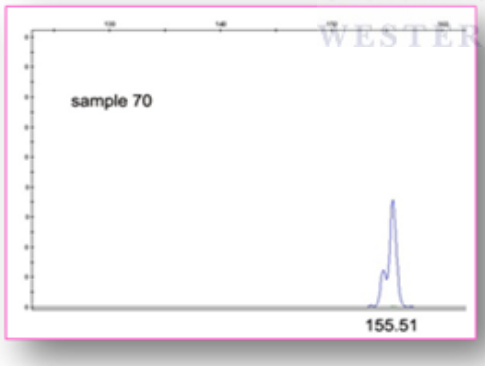
Non Informative samples continued:

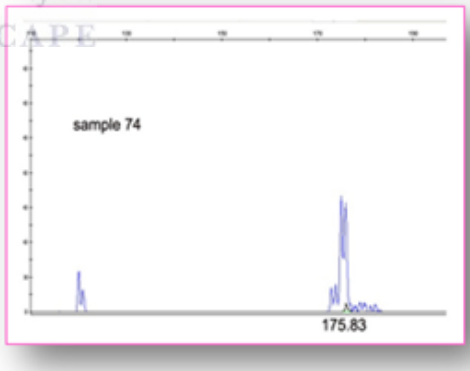












Appendix VI

Table representing Clinical data (identity of patients with held) of 101 samples:

Sample number, age, ER status, PR status, HER2/neu, Triple negative cases.

Sample no.	Age	Early stage tumor (1)	Late stage tumor (2)	ER +ve (1)	ER -ve (2)	PR +ve (1)	PR -ve (2)	Expressed or not (+ve or -ve)	HER2 (+ve) (1)	Equivoeal (2)	HER2 (-ve) (3)	TNBC (4)
1	56y	1		1			2	1	1			
2	80y		2	1			2	1			3	
3	58y		2	1			2	1	1			
4	44y	1			2		2	1		2		
5	38y		2		2		2	1			3	4
6	57y	1			2		2	1	1			
7	57y		2		2		2	1	1			
8	60y		2	1			2	1			3	
9	48y				2		2	1	1			
10	59y			1			2	1			3	
11	55y			1			2	1	1			
12	64y		2	1			2	1			3	
13	36y			1			2	1			3	
14	43y		2		2		2	1			3	4
15	44y			1			2	1			3	
16	46y				2		2	1			3	4
17	45y		2	1			2	1			3	
18	35y			1			2	1			3	
19	40y				2		2	1			3	4
20	29y			1		1		2		2		
21	76y		2	1			2	1	1			
22	44y				2		2	1	1			
23	45y	1		1			2	1			3	
24	40y		2		2		2	1			3	4
25	78y		2	1			2	1			3	
26	55y		2		2		2	1	1			
27	22y	1		1			2	1	1			
28	39y	1			2		2	1	1			
29	38y		2	1			2	1	1			
30	57y	1			2		2	1			3	4
31	51y	1			2		2	1			3	4
32	49y	1			2		2	1			3	4
33	49y	1		1			2	1			3	
34	76y	1			2		2	1		2		
35	66y	1		1			2	1			3	
36	65y	1			2		2	1	1			
37	48y	1		1			2	1	1			
38	38y		2		2		2	1	1			
39	41y	1		1		1		2	1			
40	36y	1			2		2	2			3	4

41	50y	1			2		2	1		2		
42	58y	1	2	1			2	1			3	
43	64y	1			2		2	1			3	4
44	39y	1			2		2	1			3	4
45	61y		2		2		2	1			3	4
46	65y	1			2		2	1			3	4
47	55y		2		2		2	1			3	4
48	52y		2		2		2	1		2		
49	55y		2		2		2	1			3	4
50	51y		2	1			2	1			3	
51	56y			1			2	1			3	
52	45y	1			2		2	1			3	4
53	52y	1			2		2	1			3	4
54	55y			1			2	1	1			
55	36y				2		2	1	1			
56	54y				2		2	1			3	4
57	30y			1			2	1	1			
58	48y			1			2	1			3	
59	69y			1			2	1		2		
60	53y			1			2	1			3	
61	31y			1			2	1	1			
62	31y	1		1			2	1		2		
63	43y			1			2	1		2		
64	64y		2	1			2	1			3	
65	29y			1			2	1	1			
66	69y	1			2		2	1	1			
67	40y			1			2	1		2		
68	39y	1			2		2	1			3	4
69	58y		2		2		2	1	1			
70	63y				2		2	1			3	4
71	23y				2		2	1			3	4
72	59y			1			2	1			3	
73	54y			1			2	1		2		
74	39y				2		2	1	1			
75	64y			1			2	1			3	
76	57y			1		1		2			3	
77	48y			1			2	1			3	
78	53y			1			2	1	1			
79	36y				2		2	1	1			
80	44y			1			2	1			3	
81	61y			1			2	1				
82	53y			1			2	1				
83	76y			1			2	1	1			
84	72y			1			2	1			3	
85	61y				2		2	1			3	4
86	63y			1			2	1	1			
87	69y				2		2	1			3	4
88	70y			1			2	1			3	
89	56y			1			2	1			3	
90	50y				2		2	1			3	4
91	79y			1			2	1		2		
92	37y				2		2	1			3	4

93	60y			1			2	1			3	
94	60y			1			2	1			3	
95	54y			1			2	1		2		
96	63y			1			2	1			3	
97	53y			1			2	1			3	
98	56y				2		2	1			3	4
99	73y				2		2	1			3	4
100	68y				2		2	1		2		
101	50y			1			2	1	1			

Appendix VI

Table: Representing data of 101 samples:Age,tumor stage, ER and PR (-ve or +ve)

Sample no.	Age	Young (≤ 50)	Old (>50)	Early stage tumor (1)	Late stage tumor (2)	ER Positive (1)	ER Negative (2)	PR Positive (1)	PR Negative (2)
1	56y		2	1		1			2
2	80y		2		2	1			2
3	58y		2		2	1			2
4	44y	1		1			2		2
5	38y	1			2		2		2
6	57y		2	1			2		2
7	57y		2		2		2		2
8	60y		2		2	1			2
9	48y	1					2		2
10	59y		2			1			2
11	55y		2			1			2
12	64y		2		2	1			2
13	36y	1				1			2
14	43y	1			2		2		2
15	44y	1				1			2
16	46y	1					2		2
17	45y	1			2	1			2
18	35y	1				1			2
19	40y	1					2		2
20	29y	1				1		1	
21	76y		2		2	1			2
22	44y	1					2		2
23	45y	1		1		1			2
24	40y	1			2		2		2
25	78y		2		2	1			2
26	55y		2		2		2		2
27	22y	1		1		1			2
28	39y	1		1			2		2
29	38y	1			2	1			2
30	57y		2	1			2		2
31	51y		2	1			2		2
32	49y	1		1			2		2
33	49y	1		1		1			2
34	76y		2	1			2		2
35	66y		2	1		1			2
36	65y		2	1			2		2
37	48y	1		1		1			2
38	38y	1			2		2		2
39	41y	1		1		1		1	
40	36y	1		1			2		2
41	50y	1		1			2		2
42	58y		2	1	2	1			2
43	64y		2	1			2		2

44	39y	1		1			2		2
45	61y		2		2		2		2
46	65y		2	1			2		2
47	55y		2		2		2		2
48	52y		2		2		2		2
49	55y		2		2		2		2
50	51y		2		2	1			2
51	56y		2			1			2
52	45y		2	1			2		2
53	52y	1		1			2		2
54	55y		2			1			2
55	36y		2				2		2
56	54y	1					2		2
57	30y		2			1			2
58	48y	1				1			2
59	69y	1				1			2
60	53y		2			1			2
61	31y		2			1			2
62	31y	1		1		1			2
63	43y	1				1			2
64	64y	1			2	1			2
65	29y		2			1			2
66	69y	1		1			2		2
67	40y		2			1			2
68	39y	1		1			2		2
69	58y	1			2		2		2
70	63y		2				2		2
71	23y		2				2		2
72	59y	1				1			2
73	54y		2			1			2
74	39y		2				2		2
75	64y	1				1			2
76	57y		2			1		1	
77	48y		2			1			2
78	53y	1				1			2
79	36y		2				2		2
80	44y	1				1			2
81	61y	1				1			2
82	53y		2			1			2
83	76y		2			1			2
84	72y		2			1			2
85	61y		2				2		2
86	63y		2			1			2
87	69y		2				2		2
88	70y		2			1			2
89	56y		2			1			2
90	50y		2				2		2
91	79y	1				1			2
92	37y		2				2		2
93	60y	1				1			2
94	60y		2			1			2
95	54y		2			1			2

96	63y		2			1			2
97	53y		2			1			2
98	56y		2				2		2
99	73y		2				2		2
100	68y		2				2		2
101	50y		2			1			2

Appendix VII

Table: Representing data of 101 samples: LOH(NI=non informative, ND=non detectable), gene expression of *AKT1*, *AKT2*, *PIK3*, *FGFR*, *PTEN*, HER2, ER

Sample no.	LOH NORMAL AND TUMOR	HER2	AKT1	AKT2	ER	PIK3CA	FGFR2	PTEN	Number of genes altered
1	LOH	4.89		2.17	1.71	4.89	1.31	1.15	4
2	NO LOH	-	-	-	-	-	-	-	-
3	NO LOH	-	-	-	-	-	-	-	-
4	NI	-	-	-	-	-	-	-	-
5	LOH	-	-	-	-	-	-	-	-
6	NO LOH	-	-	-	-	-	-	-	-
7	NI	-	-	-	-	-	-	-	-
8	NI	1.4	2.4	1.76	0.55	1.4	1.75	1.37	4
9	NO LOH	-	-	-	-	-	-	-	-
10	NO LOH	-	-	-	-	-	-	-	-
11	LOH	-	-	-	-	-	-	-	-
12	LOH	-	-	-	-	-	-	-	-
13	NO LOH	-	-	-	-	-	-	-	-
14	LOH	-	0.35	1.97	0.27	-	-	1.01	4
15	LOH	-	1.59	0.3	-	-	-		1
16	NO LOH	-	-	-	-	-	-	-	-
17	LOH	-	-	-	-	-	-	-	-
18	NO LOH	-	-	-	-	-	-	-	-
19	NI	-	-	-	-	-	-	-	-
20	NO LOH	-	1.5	-	-	-	-	-	0
21	NO LOH	-	-	-	-	-	-	-	-
22	NO LOH	-	-	-	-	-	-	-	-
23	ND	-	-	-	-	-	-	-	-
24	ND	-	-	-	-	-	-	-	-
25	ND	-	-	-	-	-	-	-	-
26	ND	-	-	-	-	-	-	-	-
27	ND	-	-	-	-	-	-	-	-
28	LOH	-	-	-	-	-	-	-	-
29	LOH	10	2.81	1.39	0.31	-	-	1.98	4 high
30	ND	-	-	-	-	-	-	-	-
31	NI	4	-	-	-	-	-	-	2
32	NO LOH	4	1.19	0.26	-	-	-	1.15	5
33	ND	-	-	-	-	-	-	-	-
34	ND	-	-	-	-	-	-	-	-
35	LOH	50	1.35	-	-	-	-	0.63	4 high
36	ND	-	-	-	-	-	-	-	-
37	NI	13	1.24	0.56	1.18	-	-	0.32	6 high
38	ND	-	-	-	-	-	-	-	-
39	ND	-	-	-	-	-	-	-	-
40	NI	-	-	-	-	-	-	-	-
41	LOH	2	0.06	-	-	-	-	-	2
42	LOH	2	0.37	-	-	-	-	-	2
43	NI	4		-	-	-	-	-	2

44	LOH	0.99	2.77	2.05	2.46	0.99	1.69	1.89	4
45	NO LOH	2	–	–	–	–	–	–	0
46	LOH	2	–	–	–	–	–	–	0
47	NI	2	–	–	–	–	–	–	0
48	NO LOH	2.38	1.34	2.23	0.67	2.38	1.75	1.74	3
49	NO LOH	1.42	2.02	0.54	0.60	1.42	0.53	0.4	5
50	LOH	2	1.87	1.52	0.79	–	–	0.61	2
51	LOH	2	–	–	–	–	–	–	0
52	NI	2	–	–	–	–	–	–	0
53	NO LOH	13	2.19	1.95	0.37	–	–	0.67	4 high
54	LOH	28	–	–	–	–	–	–	1
55	NO LOH	10.54	0.94	0.71	0.43	10.54	0.51	–	6 high
56	LOH	2	1.13	2.04		–	–	1.06	1
57	NI	12	1.46	1.27	0.29	–	–	0.64	6 high
58	LOH	2	–	–	–	–	–	–	1
59	LOH	2	–	–	–	–	–	–	1
60	NI	2	–	–	–	–	–	–	1
61	ND	8.94	1.37	2.02	1.23	8.94	1.31	1.83	5 high
62	NI	2.6	1.45	1.24	0.76	2.6	1.34	0.53	5
63	NI	1.28	1.62	1.4	0.69	1.28	–	0.4	6
64	NI	1.47	2.1	2.22	0.58	1.47	1.9	2.09	4
65	NI	7.95	2.25	2.72	7.21	7.95	1.78	1.58	5 high
66	NI	4.39	–	–	–	4.39	–	–	3
67	LOH	2.47	1.52	1.81	0.47	2.47	2.15	2.57	4 high
68	ND	–	–	–	–	–	–	–	–
69	NO LOH	–	–	–	–	–	–	–	1
70	NI	–	0.81	–	–	–	–	–	2
71	LOH	2.24	1.55	2.45	0.34	2.24	1.87	2.24	1
72	NI	–	2.12	–	0.30	–	–	0.67	3 high
73	NO LOH	1.26	2.31	1.49	0.69	1.26	1.27	0.77	5
74	NI	13.72	1.2	1.4	0.97	13.72	1.18	0.65	7 high
75	NO LOH	–	1.48	0.71	–	–	–	0.31	3
76	LOH	–	–	–	–	–	–	–	–
77	NO LOH	1.79	2.67	–	1.21	1.79	2.23	2.16	2
78	NO LOH	4.74	2.18	–	1.43	4.74	1.7	2.14	2
79	ND	4.94	2.43	–	2.63	4.94	1.98	3.1	5
80	NI	2.44	2.23	1.87	0.85	2.44	1.84	1.86	1
81	NI	–	2.9	1.24	1.19	–	2.19	1.17	5 high
82	ND	3.01	1.16	1.63	1.97	–	1.69	2.2	3
83	ND	12.52	2.66	1.69	0.19	12.52	–	2.96	6 high
84	ND	–	1.64	–	–	–	–		0
85	NO LOH	1.07	1.5	1.78		1.07	–	1.29	4
86	ND	1.79	2.36	2.78	0.60	1.79	–	1.75	3
87	ND	–	–	–	0.00	–	–		0
88	ND	–	–	–	–	–	–		–
89	ND	2.14	2.12	2.73	0.77	2.14	–	1.68	1
90	ND	1	1.65	2.45	1.03	1	–	1.63	3
91	ND	2.66	2.48	1.9	1.01	2.66	2.04	1.84	3
92	ND	–	–	–	–	–	–	–	–
93	ND	–	–	–	–	–	–	–	–
94	ND	2.12	2.07	0.8	0.93	2.12	–	2.29	3
95	ND	–	–	–	–	–	–	–	–

96	ND	–	–	–	–	–	–	–	–
97	ND	–	0.75	1.37	0.6	–	–	0.97	5
98	ND	–	0.57	1.2	–	–	–	0.66	4
99	ND	–	2.92	–	–	–	–	0.84	3
100	ND	–	–	–	–	–	–	–	1
101	ND	–	2	1.11	–	–	0.45	1.35	4

UNIVERSITY OF CAPE TOWN

Health Sciences Faculty
Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6626 • **Facsimile** [021] 406 6411
e-mail: shuretta.thomas@uct.ac.za

01 October 2010

HREC REF: 454/2010

Prof D Govender & Dr SH Abdul-Rasool
Anatomical Pathology

Dear Prof Govender & Dr Abdul-Rasool

PROJECT TITLE: DETECTION OF TUMOUR DNA IN BLOOD AND PRIMARY TUMOURS OF BREAST CANCER PATIENTS: CLINICAL AND PROGNOSTIC CORRELATION.

Thank you for your letter to the Faculty of Health Sciences Human Research Ethics Committee dated 28th September 2010.

It is a pleasure to inform you that the Ethics Committee has only **granted approval** for the use of Archival slides that have been de-identified from the NHLS Services in Cape Town at the Division of Anatomical Pathology at UCT.

Approval is granted for one year till the 15th October 2011.

Please submit an annual progress report if the research continues beyond the approval period. Please submit a brief summary of findings if you complete the study within the approval period so that we can close our file.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the REC. REF in all your correspondence.

Yours sincerely

Wesley Blockman

PP **PROFESSOR M BLOCKMAN**
CHAIRPERSON, HSF HUMAN ETHICS
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

S Thomas

This serves to confirm that the University of Cape Town Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

S Thomas