UNIVERSITY of the
WESTERN CAPE

DST/MINTEK
NIC
Nanotechnology Innovation Centre

Identification of microRNAs as a class of biomarkers for the early diagnosis of prostate cancer: an *in silico* and molecular approach.

Chipampe Patricia Lombe

Student Number: 3111206

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

Supervisor: Dr Ashley Pretorius

Co-supervisor: Professor Mervin Meyer

13 November 2015

Bioinformatics Research Group

Keywords

Prostate cancer,

miRNA,

biomarkers,

early diagnostics,

qRT-PCR,

bioinformatics,

Expression profiling

in silico.

Abstract

Prostate cancer (PCa) is the second most common form of cancer in men around the world. In many parts of Africa, data on prostate cancer is sparse. This is attributed to poor access to testing and diagnostics. The International Agency for Research on Cancer (GLOBOCAN) estimated that 28,000 deaths occurred as a result of PCa in Africa in 2008, 4500 of which were in South Africa. This figure (28,000) is predicated a rise to 57,000 over the next two decades. Currently, the most commonly used diagnostic tests for PCa are the DRE and PSA tests. The former is highly invasive and both have low specificity and poor sensitivity. Therefore, the need for a less invasive early detection method with the ability to overcome the lack of specificity and sensitivity is required. Biomarkers have recently been identified as a viable option for early detection of disease. Examples of biological indicators for disease are miRNAs. miRNAs are small non-coding RNA molecules which play a key role in controlling gene expression and certain biological processes. Studies have shown that aberrantly expressed miRNAs are a hallmark of several diseases like cancer. miRNA expression has been shown to be associated with tumour development, progression and response to therapy, suggesting their possible use as diagnostic, prognostic and predictive biomarkers.

The study aimed to investigate the potential of miRNAs implicated in prostate cancer as putative biomarkers for the disease and evaluating these miRNAs in a panel of prostate as well as several other cancer cell lines using qRT-PCR. An *in silico* approach was used to identify 13 putative miRNAs implicated in prostate cancer of which 8 were further analysed in a parallel study and 5 in this study.

Two publicly available target prediction software were used for target gene prediction of the 5 identified miRNAs. The target genes were subjected to functional analysis using web-based software, DAVID. Functions which were clustered with an enrichment score of 1.3 and greater were considered significant. Targets with gene ontologies linked to “transcription regulation”, “regulation of “apopotosis”, “extracellular region” and “metal ion binding” were considered for further analyses. Protein gene interaction analysis was performed to determine the pathways the target genes are involved in using STRING. Expression profile analysis of the genes in various tissues was also carried out using *in silico* methods through the TiGER and GeneHub-GEPIS databases.

Analysis using DAVID resulted in 9 gene targets for the 5 miRNAs. It was found that miR3 seemed the most promising miRNA for biomarker validation based on the *in silico* analyses. Its target gene MNT was found to be abundantly expressed in prostate tissue from the TiGER results. The GeneHub-GEPIS results also indicated that the gene’s expression is up-regulated during prostate cancer.

The expression levels of the miRNAs analysed using qRT-PCR indicated that miR3 is significantly over-expressed in prostate cancer cells when compared to the other cancer cell lines used in this study, corroborating the results observed from the *in silico* analyses. Another miRNA with interesting results was miR5. It was predicted to target two genes, YWHAZ and TNFSF13B. In TiGER, both were found to be expressed in prostate tissue. The genes were also found to be up regulated during prostate cancer in GeneHub-GEPIS. The expression level of miR5 in LNCaP was 15.32; it was significantly up-regulated in the cell line using qRT-PCR. However, miR5 was also present in HEPG2-7.06, MCF7-0.79, HT29-

1.61 and H157-3.59. Thus, it was concluded it can be used as a biomarker in combination with other miRNAs. The miRNA miR2 was found to target the actin filament protein encoding gene AFAP1. The gene was predicted to be up-regulated with a DEU of 33.25 in GeneHuB-GEPIS. The qRT-PCR analysis showed that the expression ratio in LNcaP was 8.79. However, miR2 expression was up-regulated in MCF7-0.85 and HT29-1.09 as well. The expression level of miR1 in BHP1 was found to be 4.85. It can be considered as an indicator for benign prostate hyperplasia.

Future work would include investigating the expression of miR3 in a larger panel of cancer cells as well as in patient samples. In addition, analysis of the UTR sequences of the miRNAs targets experimentally to prove that the target genes identified using *in silico* methods, are indeed regulated by these miRNAs. Furthermore, performing gene "knock-out" studies on the genes that code for the miRNAs to study their roles in prostate cancer development.

November, 2015

Declaration

I declare that Identification of microRNAs as a class of biomarkers for the early diagnosis of prostate cancer: an *in silico* and Molecular approach is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name..... Date.....

Dedication

For my father, Francis Kangwa Lombe

Acknowledgements

First and foremost, I am grateful to the Almighty for constantly blessing me with the wisdom, courage and wellbeing needed to complete this project. My aunt and my mother, Esther Jordan, for your encouragement and support in the past 20 years and even more these past two years.

My sincere gratitude goes to Dr Ashley Pretorius for allowing me the opportunity to work with the Bioinformatics Research Group. Thank you for always having your door open and sharing your valuable expertise. You are the first lecturer I met at UWC and since then, you have been very patient and encouraging. THANK YOU! I could not have asked for a better supervisor.

Professor Mervin Meyer, thank you for always having your door open, thank you for your patience and kindness and most importantly, for helping to make my studies at UWC possible by facilitating the DST/MINTEK bursary.

Thank you to the members of the BRG, Habeeb, Marius and Yancke, thank you for always being there with advice when I was lost with my writing. Mustafa and Nicole, thank you for all your advice with lab work and tissue culture. Namhla, thank you for your help with the qRT-PCR result analysis. Firdous, thank you for all the preliminary work you did on the project. Fatima, Sekuru and Riziki, you always knew how to lift my spirits throughout the year, thank you.

My friends Tiza, and Bethsheba (my Dada), you have been pillars of strength and immeasurable support throughout this year, I thank you. To the members of the Biotechnology department, I can't name you all, thank you for everything you have shared with me this year. Thank you to the University of the Western Cape

for allowing me to complete this part of my education at the institution. To my brother Phillip and my Vatika, thank you.

List of Abbreviations

/min	per minute
µL	microlitres
A549	Human alveolar basal epithelial cells
ACS	American Cancer Society
AFAP1	Actin Filament Associated Protein
AJCC	American Joint Committee on Cancer
AUA	American Urological Association
BIRC2	Apototic suppressor
BPH	Benign Prostatic Hyperplasia
BPH1	Benign Prostatic Hyperplasia 1 cells
CANSA	Cancer association of South Africa
CDC	Centers for Disease Control and Prevention
cDNA	Complementary Deoxyribonucleic acid
CRM	Cis-regulatory module
CTNND1	(E-cadherins) cell maligancy
CZ	central zone
DAVID	**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery
DEU	**D**igital **E**xpression **U**nit
DNA	Deoxyribonucleic acid
DRE	Digital rectal examination
EST	Expressed sequence tag
FOXC1	Forkhead box C1

g	gravitational force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GLOBOCAN	The International Agency for Research on Cancer
HeLa	Cervical adenocarcinoma cells
HEPG2	Hepatocellular carcinoma cells
HT29	Human Colon adenocarcinoma cells
LIG	Ligase IV, DNA, ATP-dependent
LNCaP	Lymph Node Prostate adenocarcinoma
MCF12	Human mammary epithelial cells 12
MCF7	Human Breast Adenocarcinoma 7 cells
mg	milligrams
miRNA	Micro Ribonucleic acid
mL	millilitres
MNT	MAX binding protein
mRNA	Messenger Ribonucleic acid
NA	Nucleic Acid
NCBI	National Centre for Biotechnology Information
NCI	National Cancer Institute
NIH	National Institutes of Health
PCa	Prostate Cancer
PIN	Prostatic intraepithelial neoplasia
PNT1a	Prostate Normal Epithelium cells
PSA	Prostate-Specific Antigen
qRT-PCR	Quantitative real time polymerase chain reaction

RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
s	seconds
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TiGER	Tissue-specific Gene Expression and Regulation
TF	Transcription factor
TNFSF13B	Tumour Necrosis Factor Superfamily 13B
TNFSF15	Tumour Necrosis Factor (ligand) Superfamily15
TNM	Staging Primary tumour lymph node metastasis
TRUS	transrectal ultrasound
TZ	transition zone
USPSTF	United States Preventive Services Task Force
YWHAZ	Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein, zeta polypeptide
WHO	World Health Organization

List of Figures

Chapter 1

Figure 2.8: Expression profile for MNT from TiGER (A) MNT is preferentially expressed in the prostate with an expression enrichment value greater than 2. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.9: Expression profile for YWHAZ from TiGER (A) YWHAZ is preferentially expressed in the tongue with an expression enrichment value greater than 2.5. Expression in prostate tissue is between 1 and 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.10: Expression profile for LIG4 from TiGER (A) LIG4 is preferentially expressed in muscle tissue with an expression enrichment value of about 4.5. There is no expression recorded in prostate tissue. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.11: Expression profile for FOXC1 from TiGER (A) FOXC1 is preferentially expressed in the heart with an expression enrichment value greater than 6. Expression in prostate tissue is between 3 and 4. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.12: Expression profile for TNFSF13B from TiGER (A) TNFSF13B is preferentially expressed in the thymus with an expression enrichment value

greater than 13. Expression in prostate tissue is about 2.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.13: Expression profile for BIRC2 from TiGER (A) BIRC2 is preferentially expressed in the thymus with an expression enrichment value greater than 4.5. Expression in prostate tissue is less than 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.14: Expression profile for CTNND1 from TiGER (A) CTNND1 is preferentially expressed in the bladder with an expression enrichment value greater than 2.5. Expression in prostate tissue is between 1 and 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

Figure 2.15: Expression profile for TNFSF15 from GeneHUB-GEPIS. Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow. There was no expression profile available for the gene in the TiGER database.

Figure 2.16: Expression profile for AFAP1 from GeneHUB-GEPIS. Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow. There was no expression profile available for the gene in the TiGER database.

Chapter 3

List of equations

Chapter 3

List of tables

Chapter 1

Chapter 2

Chapter 3

Table of Contents

Chapter 1

Literature Review

1.1 Cancer

One of the greatest health concerns worldwide currently is cancer. It has become the most devastating disease among the chronic diseases. It has continued to rise in incidence and is associated with a high morbidity and mortality rate, in 2012; there were 8.2 million cancer related deaths (CDC, 2014). According to the National Cancer Institute statistics, there are over 13 million new cases of cancer per annum worldwide, and the number of new cases is expected to rise by 70 % by 2050 with deaths projected to rise to over 12 million worldwide by 2040 (WHO, 2015). Additionally, the Cancer Association of South Africa reports that one in six males and one in seven females have some type of cancer with prostate cancer being most prevalent in the former and breast cancer in the latter.

Cancer arises in almost all human tissues and it is believed that the basic processes that transform a normal cell into a cancer cell are essentially the same in all cancers arising in the human body. These basic properties of survival, proliferation and dissemination are called the hallmarks of cancer (Hanahan and Weinberg, 2000). Although these hallmarks are thought to be common for all types of cancer, they are acquired through diverse distinct mechanisms during different times of the multistep tumourigenesis process in different forms of cancer. There are six hallmarks of cancer(figure 1) originally proposed by

Hanahan and Weinberg, these are: i) sustaining proliferative signalling, ii) evading growth suppressors, iii) activating invasion and metastasis, iv) enabling replicative immortality, v) angiogenesis and vi) resisting cell death (Hanahan and Weinberg, 2011).

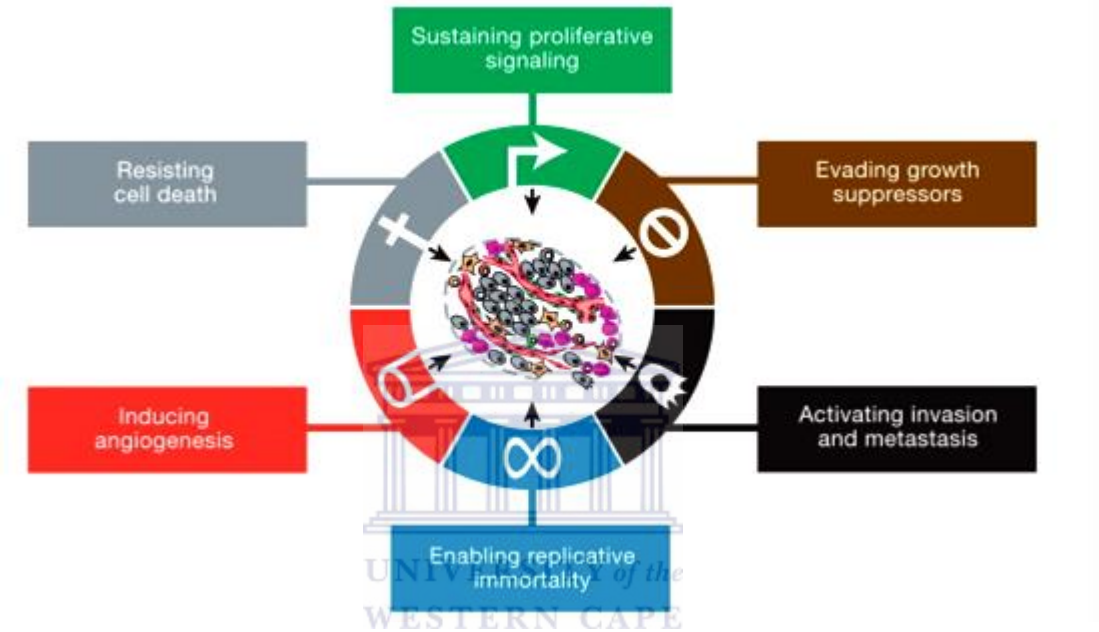


Figure 1: The six hallmarks of cancer acquired by normal cells as they evolve to a neoplastic state (Adapted from Hanahan and Weinberg, 2011).

1.1.1 Hallmarks of cancer

The hallmarks of cancer have been defined as acquired functional capabilities that allow cancer cells to survive, proliferate, and disseminate (Hanahan and Weinberg 2000; Hanahan and Weinberg 2011). These functions are acquired in different tumour types via distinct mechanisms and at various times during the

course of multistep tumourigenesis (Hanahan and Weinberg 2000; Colotta *et al*., 2009). Recently, an increasing body of research has suggested that two additional hallmarks of cancer are involved in the pathogenesis of some and perhaps all cancers. One involves the capability to modify, or reprogram, cellular metabolism in order to most effectively support neoplastic proliferation (Hanahan and Weinberg 2011; Colotta *et al*., 2009; Luo *et al*., 2009). The second allows cancer cells to evade immunological destruction, in particular by T and B lymphocytes, macrophages, and natural killer cells (Hanahan and Weinberg 2011; Colotta *et al*., 2009; Negrini *et al*., 2010). Because neither capability is yet generalized and fully validated, they are referred to as emerging hallmarks (Hanahan and Weinberg, 2011).

Additionally, two consequential characteristics of neoplasia facilitate acquisition of both core and emerging hallmarks (figure 1.2). Genomic instability and thus mutability present cancer cells with genetic alterations that drive tumour progression (Colotta *et al*., 2009; Luo *et al*., 2009). Inflammation by innate immune cells designed to fight infections and heal wounds can instead result in their inadvertent support of multiple hallmark capabilities, thereby manifesting the now widely appreciated tumour-promoting consequences of inflammatory responses (Hanahan and Weinberg 2011; Colotta *et al*., 2009; Luo *et al*., 2009; Negrini *et al*., 2010).

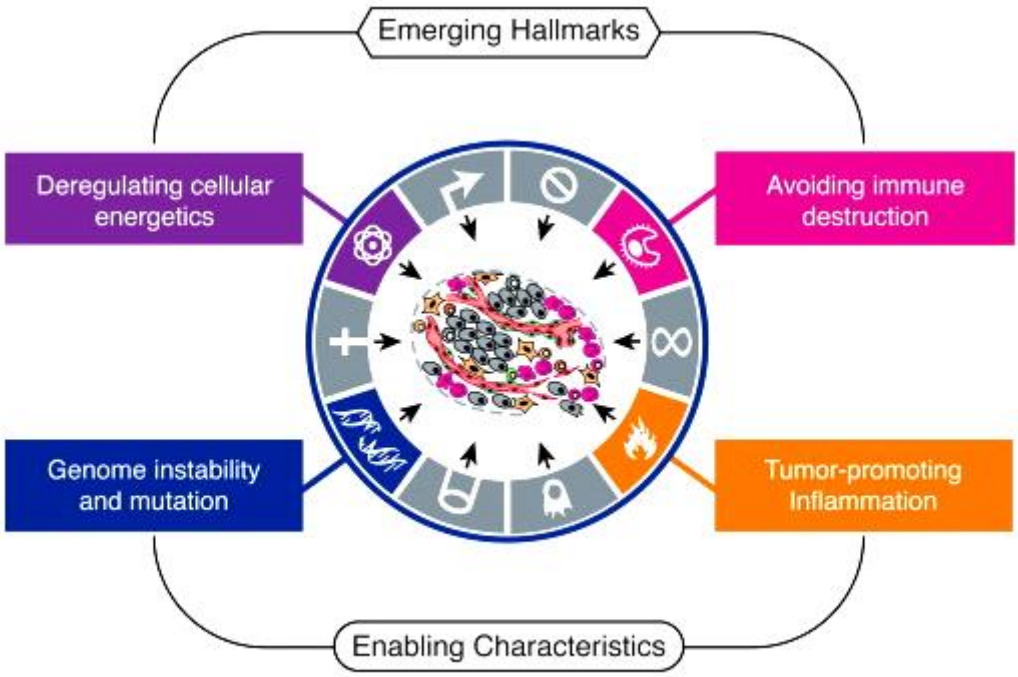


Figure 1.2: Emerging Hallmarks of cancer and Enabling Characteristics. (Adapted from Hanahan and Weinberg, 2011).

1.2 Prostate cancer (PCa)

Prostate cancer is one of the most common cancers among men worldwide. According to the International Agency for Research on Cancer (GLOBOCAN) in 2012, prostate cancer was among the most commonly diagnosed cancers in males, coming second after lung cancer. Prostate cancer makes up 8 % of all cancers diagnosed in the world for both sexes (figure 1.3).

The incidence of PCa varies significantly between ethnic groups and geographic regions with men of African descent living outside of Africa having some of the highest incidence rates, 234.6 per 100,000 (Chang *et al*., 2011; Rebbeck *et al*., 2013). In the USA, men of African descent are more frequently diagnosed with PCa at an earlier age of onset, have higher prostate specific antigen levels, higher tumour volume as well as more aggressive tumour stages (Zeeger *et al*., 2004; Lange *et al*., 2008).

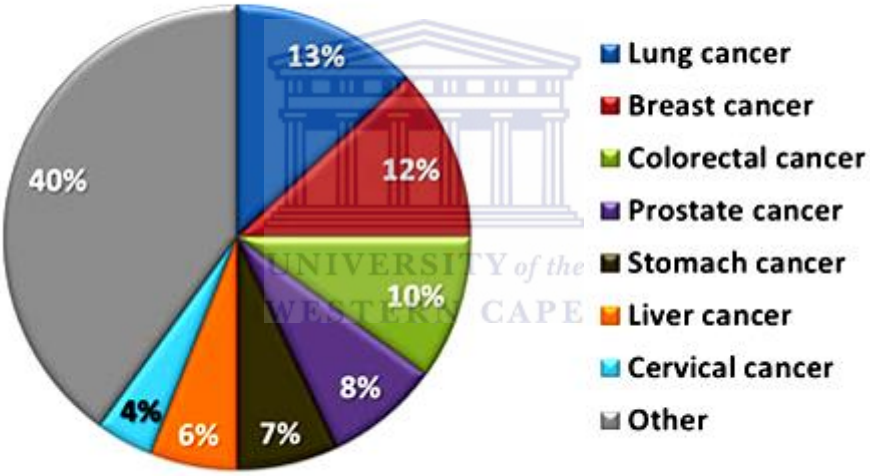


Figure 1.3: The most commonly diagnosed cancers worldwide (for males and females) in 2012, with prostate cancer accounting for 8 %. (Adapted from GLOBOCAN, 2012).

Mortality rate among people of African descent is also higher than that of Caucasian Americans with the latter having 25.6 deaths per 100,000 men and the former having 2.4 times that rate at 62.3 per 100,000 men (Chang *et al*., 2011). In

contrast, prostate cancer incidence in Asian countries is low (Curado *et al*., 2007) and reasons for this are not clear. However, studies do indicate that these patterns may be linked to differences in access to care, inherited susceptibility, referral patterns, differences in the biology of the disease, treatment options, differences in testing, reporting and diagnosis of the disease. (Parkin *et al*., 2003; Rebbeck *et al*., 2013).

In many parts of Africa, data on prostate cancer is sparse. This could be because of poor access to testing and diagnostics. The International Agency for Research on Cancer (GLOBOCAN) estimated that 28,000 deaths occurred in Africa in 2008 and predicated a rise to 57,000 over the next two decades. However, a study by Chu *et al.*,2011 questioned the validity of these figures, stating that they could be an underestimation of PCa cases in Africa. In their study on PCa among different racial groups in the Western Cape, Heyns *et al*., 2011, suggestes the underestimates could be as a result of less awareness and education about the disease among patients and physicians as well as the fear and taboo of possible digital rectal examinations.

1.2.1 Anatomy and normal functions of the prostate

The prostate is a compound exocrine gland that is part of the male reproductive system (Figure 1.4) (Theodorescu, 2001). The normal adult prostate is about the size of a walnut and increases in size with age. It is situated at the base of the bladder and surrounds the urethra. The rectum sits posterior to the prostate, allowing for palpation of the prostate during rectal examination (Basch *et al*., 2012). The secretions of the prostate are high in sugars and proteins. The prostatic secretions are thought to be important in aiding fertilization by increasing the motility of sperm and perhaps promoting the viability of sperm after ejaculation. The gland produces approximately 20 % of the fluid produced during ejaculation (Gupta *et al*., 2013). The glandular tissue of the prostate is dependent on androgens for normal growth and development; androgens are hormones that promote the development of male sexual characteristics, the most common being testosterone (Theodorescu, 2001).

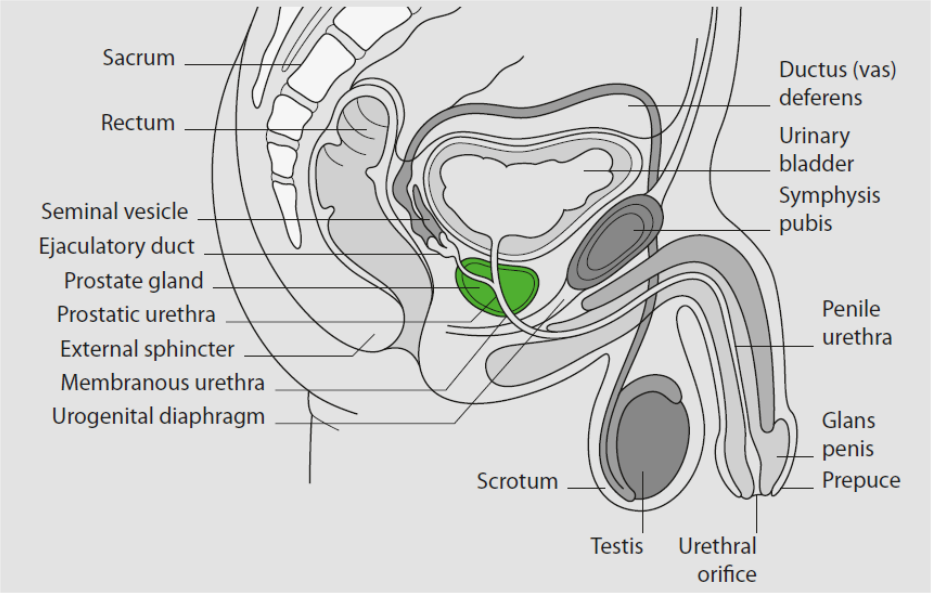


Figure 1.4: Anatomy of the prostate gland. (Adapted from Theodorescu, 2001).

In terms of structure, the prostate is clinically divided into zones (figure 1.4). These are the peripheral zone (PZ), which comprises of the posterior part of the gland surrounding the distal urethra and it has been reported that 80 % of cancers arise in the PZ (Taylor *et al*., 2011). The central zone (CZ) surrounds the ejaculatory ducts and approximately 5 % of cancers are reported to arise here (Descotes *et al*., 2007). Figure 1.3 also shows that the transition zone (TZ) surrounds the proximal urethra and continues to enlarge throughout life and is the part of the gland where benign prostatic hyperplasia (BPH) occurs in later life. About 15 % of cancers originate from the transition zone (Taylor *et al*., 2011).

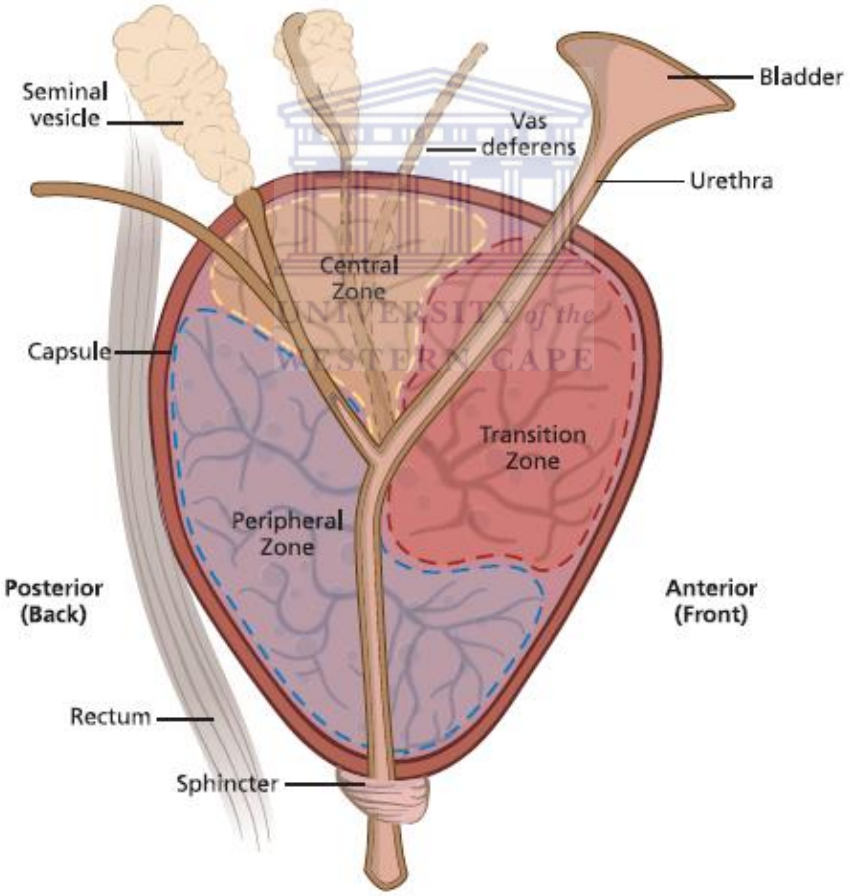


Figure 1.5: Zonal anatomy of the prostate. The human prostate is composed of zones of exocrine glandular tissue with the ducts emptying into the prostatic urethra. (Adapted from Cramer, 2007).

1.2.2 Prostate cancer pathology

The prostate has two main epithelial cell types; the basal cells and the luminal cells (Theodorescu *et al*., 2001). The latter lines the lumen of the prostate and is responsible for the production of a protein known as the prostate specific antigen (PSA). There has been no record of any other cell in the male body producing this protein and thus it is used as a biomarker for prostate cells. The basal cells surround the luminal cells and act as a barrier between the luminal cells and the stroma (Taylor and Albertsen, 2011). During the development of prostate cancer, the normal prostate structure is altered, the rate of cell division surpasses cell death and this leads to uncontrolled tumour growth. These changes result in a breakdown of the basal cell barrier between the prostatic duct and the surrounding stroma. These breakdowns lead to an invasion of luminal cells into the surrounding stroma, which can eventually lead to migration of these cells into the rest of the body using the nervous or circulatory systems (Huang *et al*., 2013). Arriving at their final destination, the tumour cells can lodge and grow secondary tumours which can result in a dramatic decline in the cure rates for the disease. The presence of these prostate cancer cells in another site, such as bone, does not change its classification to bone cancer-for instance. The new tumour is still considered to be prostate cancer (Taylor and Albertsen, 2011).

1.2.3 Benign Prostatic Hyperplasia

Benign Prostatic Hyperplasia (BPH) arises from the glandular epithelial cells that line the ducts of the prostate gland. Although the pathogenesis of BPH is not well understood, it is generally agreed that it begins with stromal alterations, which then stimulate growth and variably alter the differentiation of associated

epithelial cells (Taylor and Albertsen, 2011). It is interesting to note that most prostate cancers arise in prostates that already have BPH. However, BPH originates in the transitional zone while the peripheral zone is the most prevalent site for prostate cancer. BPH is easily distinguished from prostate cancer histologically, as BPH has a distinct basal cell layer and is characterised by an altered stromal-epithelial arrangement (Descotes *et al*., 2007).

BPH is the most common non-malignant condition to affect men and is the most frequent benign condition found in the prostate, occurring in more than 70 % of men aged 70 years or older (Taylor *et al*., 2011). Although BPH is not the premalignant precursor of prostate cancer (Basch *et al*., 2012), this condition has significant similarities with prostate cancer. Both show increased prevalence with age (although BPH usually occurs 15-20 years earlier), they both require androgenic stimulation and may respond to androgen deprivation (Botswick and Brawer, 1987). The current clinical marker for prostate cancer, PSA, is known to increase with age and is also associated with BPH (Gupta et al., 2013). Thus, there may be considerable uncertainty when distinguishing between BPH and prostate cancer using the PSA test (Basch *et al*., 2012).

1.2.4 Prostatic intraepithelial neoplasia: A precursor to prostate cancer

The piling up of luminal cells in the prostate is called Prostatic Intraepithelial Neoplasia (PIN). PIN is thought to bea precursor to prostate cancer and is usually

segregated into low grade and highgrade, depending on how closely it resembles true prostatecancer (Taylor and Albertsen, 2011). Prostatic intraepithelial neoplasia retains a basal cell layer, distinguishing it from prostate cancer, which lacks a basal celllayer. In figure 1.6, the normal luminal cells are very similar to one another in size, shape, and the location of the nucleus toward the basal layer (Botswick and Brawer, 1987). In low-grade PIN the luminal cells become less uniform; the nuclei are no longer located solely at the basal layer, instead they become enlarged and contain large dark spots referred to as nucleoli. In high-grade PIN these characteristics become more pronounced, the nuclei become very large and the nucleoli are very prominent. In early prostate cancer (carcinoma), there is an additional loss of the complete basal layer.

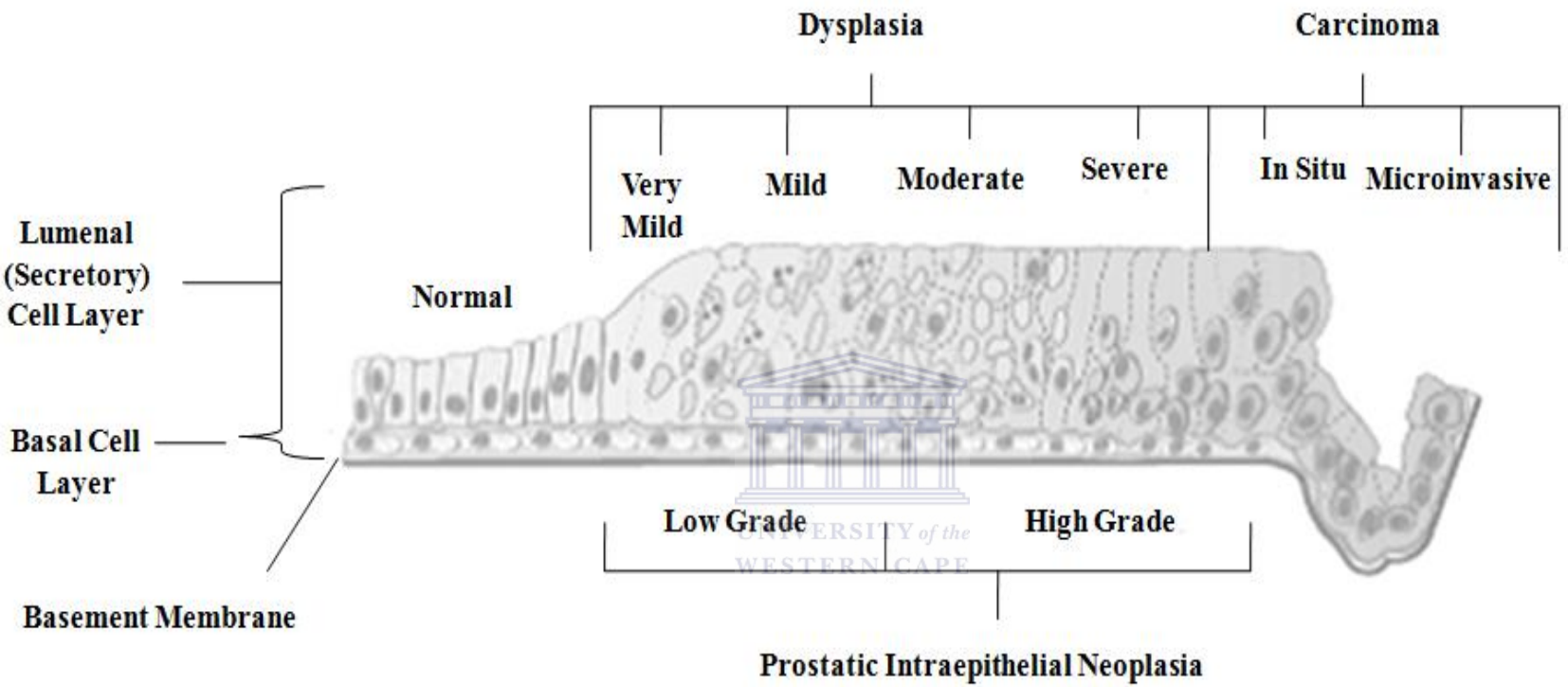


Figure 1.6: Cellular changes in Prostatic Intraepithelial Neoplasia (PIN). The progressive changes in the structure of the gland are seen going from left (normal) to right (microinvasive carcinoma). (Adapted from Boswick and Brawer, 1987).

1.3 The Gleason grading system

One of the defining attributes of prostate cancer that distinguishes it from PIN and the normal prostate glands is the loss of the basal layer of cells (Boswick and Brawer, 1987). However, loss of the basal layer is not the end of the process of changes in the prostate that lead to aggressive prostate cancer. Just as with PIN, in which there are low-grade and high-grade classifications of aggressiveness, there are specific changes that occur in the prostate that have been associated with more or less aggressive disease (Taylor and Albertsen, 2011). There are several systems that have been developed to classify these changes and correlate them with patient survival and response to therapy. The system that is used throughout the world is the Gleason grading system developed by Donald Gleason in the 1960s (Oesterling and Moyad, 1997; American Joint Committee on Cancer, 2010). It is a standardized system of grading histological changes in the prostate glandular structure that could be used to predict prostatic disease progression (Gleason, 1966). Figure 1.7 shows a figure that Gleason published in 1966 that is still used by pathologists to grade prostate cancer using the Gleason system. The figure represents the appearance of the prostate glandular morphology at low magnification using a microscope.

In figure 1.7, looking at panel 1, each circle represents the lumen of a duct or gland. Gleason noticed that as the prostate cancer becomes more aggressive the glandular structure becomes less uniform and more disorganized. This organization is often referred to as the degree of differentiation. The more organized (grade 1), the more differentiated. The less organized the less

differentiation and the more aggressive the cancer (grade 5) (Gleason, 1966). From 1 to 5, each number represents an increasing Gleason grade. There are a variety of glandular types that can be classified as belonging to each specific grade. As the Gleason grade increases, the size and shape of the glands become less uniform. In very advanced cancer, Gleason grade 4 or 5, the glands become very tiny, or in some cases there is a loss of glandular structure completely (Oesterling and Moyad, 1997).

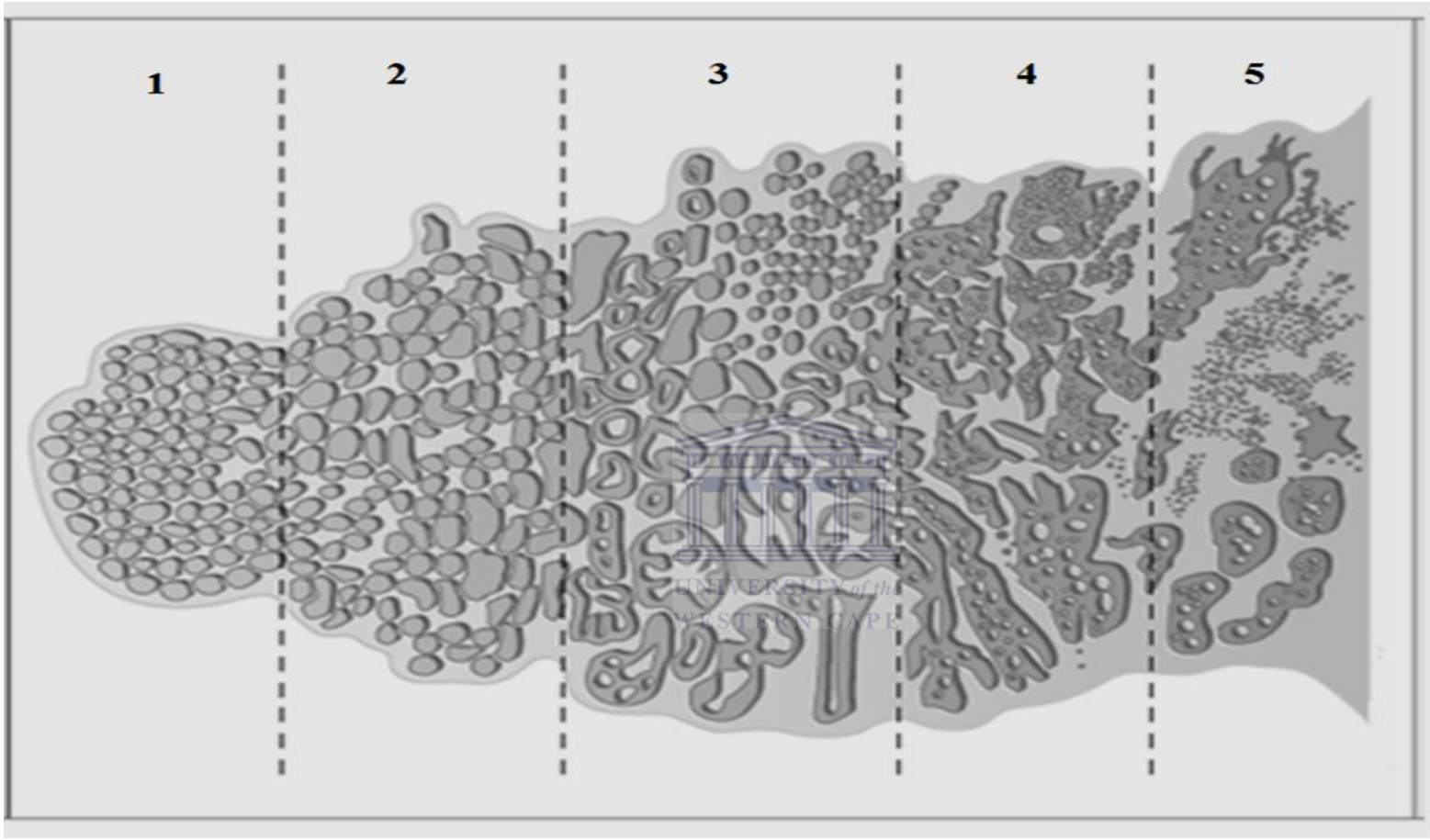


Figure 1.7: Gleason Grading System. As prostate cancer becomes more aggressive, the glands become less organized, withsmaller and more variable lumen sizes. Each panel from 1 to 5 represents an increasing Gleason grade. (Adapted from Foster and Bostwick, 1998).

1.3.1 Tumour staging

Whilst grading refers to how aggressive a cancer can be, cancer staging is the process of determining whether the cancer has spread and how far. Staging also provides a better insight into the risk of the disease spreading further so that the correct treatment option can be selected. The TNM stage was developed by the American Joint Committee on Cancer (AJCC) (Wallace *et al*., 1975). It is used to evaluate the extent of the primary tumour (**T**), the affected regional lymph nodes (**N**) and if it has spread or metastasized (**M**) table 1.

Table 1: TNM staging of prostate cancer.

Primary Tumour (T)	**TX**	Cannot evaluate the primary tumour
	T0	No evidence of primary tumour
	Ta	Non invasive papillary carcinoma
	T	Carcinoma in situ: flat tumour
	T2	Tumour invades muscle **T2a** Tumour invades superficial muscle (inner half)
	T3	**T2b** Tumour invades deep muscle (outer half) Tumour invades perivesical tissue **T3a** Microscopically **T3b** Macroscopically
	T4	Tumour invades prostate **T4a** Tumour invades prostate **T4b** Tumour invades pelvic or abdominal wall
Regional Lymph Nodes (N)	**NX**	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Metastasis in a single lymph node
Distant Metastasis (M)	**MX**	Distant metastatic cannot be assessed
	M0	No distant metastatic
	M1	Distant metastatic

M, metastasis; N, nodes; T, tumour. (Adapted from American Joint Committee on Cancer, 2013).

1.4 Diagnosis

Diagnosis of prostate cancer includes testing for the presence of prostate-specific antigen (PSA) in blood testing, PSA velocity (how much a patient's PSA levels increase from year to year), digital rectal examination (DRE), blood count and biochemical profiling, transrectal ultrasound (TRUS) and biopsy (Horwich *et al*., 2010). Secondary diagnosis can occur during investigation for other diseases, such as bladder cancer, where biopsies may be undertaken.

1.4.1 Prostate-specific antigen testing

Currently, prostate cancer screening is done in part through the use of the prostate specific antigen (PSA) blood test often combined with a digital rectal exam (DRE). Not all low or high PSA levels are indicative of prostate cancer, as PSA levels are organ specific and not cancer specific (Descotes *et al*., 2007; Velonas *et al*., 2013; Qu *et al*., 2014). Despite routine application of PSA assays, PSA screening has been very controversial. As of October 2011, the United States Preventive Services Task Force (USPSTF) recommended against the use of PSA as a screening tool for prostate cancer in asymptomatic men. The controversy around use of PSA as a screening tool stems from the fact that PSA is organ-specific and not disease-specific, thus making it prone to high false-positive diagnosis. In addition to prostate cancer, there are several reasons for elevated levels of PSA found in a man's blood. These include: benign prostate hyperplasia, prostatitis (Basch *et al*., 2012), recent ejaculation, digital rectal exam, and prostate biopsy. Biopsies show that over two-thirds of men with PSA levels greater than 4 ng/mL, do not have prostate cancer (Descortes *et al*., 2007). In the

meantime, there are men with PSA levels in the normal range (below 4ng/mL) who have prostate cancer (Heidenreich *et al*., 2011). According to the Mayo Clinic, 76 % of men with raised PSA levels do not test positive for it upon biopsy (Tollefson, 2012). Additionally, in 2004, Thompson *et al* conducted a study over a period of 7 years in 2,950 males who had never had PSA levels higher than 4.0 ng/mL or an abnormal DRE. Prostate biopsies showed that there was a 15.2 % (n=449) prevalence of PCa in men with PSA levels no higher than 4.0 ng/mL. High-grade prostate cancer (defined as Gleason score ≥7) was also seen in 15.8 % (n = 71) of these men.

In 2006, Dyche *et al*., investigated the prevalence and outcome of PSA testing for prostate cancer screening or diagnosis in men 45 years to 75 years of age over a period of 6 years. The study was conducted on 8797 males and a total of 82,672 visits were made over the time period. The findings were that 5.7 % of these men underwent at least one PSA test. Of that 5.7 %, 3.4 % were under the age of 55. Overall, the prevalence of PSA testing was 14.9 % in the 45 to 54 years old age group and 11.8 % in the 55 to 64 years old age group and 10.3 % in the 65 to 75 years old age group. The study concluded that PSA testing for prostate cancer screening and diagnosis appears to decrease with advancing age. The study was however, not conclusive on why this happens. It could be because of the problems associated with PSA discussed afore mentioned.

Another important limitation of PSA as a biomarker is its inability to identify patients with aggressive and lethal forms of prostate cancer. Because many forms of prostate cancer are apathetic and do not progress to metastasis and death, it

would be important for new biomarkers to be able to distinguish those from aggressive prostate cancer (Heidenreich *et al*., 2011). Over the last 25 years, no new blood test, genetic test or medical x-ray have been able to replace PSA. So we are in a midst of a biomarker crisis as lack of specificity of PSA requires supplementation in order to improve patient management, and to differentiate cancer from benign diseases of the prostate.

1.4.2 Digital rectal examination (DRE)

Digital rectal examinations are performed by a clinician physically examining the prostate via the rectum for any bumps or enlargements. A normal prostate is about the size of a walnut and uniformly soft and pliable. Areas of prostate cancer will often feel harder than the surrounding normal area. The hardened part is called the nodule (Heidenreich *et al*., 2011). The prostate can also be larger than normal; however this is due to BPH rather than prostate cancer. About 18 % of patients with prostate cancer can be diagnosed by a DRE regardless of PSA levels (Carter *et al*., 2013). A DRE can indicate whether a prostate biopsy is recommended for a patient, especially in more aggressive cases (Horwich *et al*., 2010). However, like PSA testing, DREs are not absolutely conclusive. DREs are dependent on the ability of a doctor to feel the differences between a normal prostate and a tumour. If the cancer is located away from the rectal surface or if it is not particularly hard, then it might be missed. Because of such problems, the American Urological Association (AUA) 2013 guidelines could not find evidence to support the continued use of DREs for first-line screening due to its lack of

sensitivity and the high possibility of missing early prostate cancer tumours, which may not be felt during the examination (Horwich *et al*., 2010).

The National Health Institute (NIH) in the UK release statistics on digital rectal examinations for the year 2014. In the report, it is indicated in a test group of men aged between 50 and 70 years old only 47.6 % underwent a DRE (Federman *et al*., 2014). An additional 6.9 % were offered a DRE but declined. The study concluded that screening for prostate cancer remains controversial. Nevertheless, the DREs are still the standard test used for prostate cancer in conjunction with other diagnostic tests such as PSA.

1.4.3 Transrectal ultrasound (TRUS)

A Transrectal ultrasound is performed by inserting a small probe in the patient's rectum; this probe emits sound waves into the patient's prostate that echo back to the probe to create video images of the prostate (Basch *et al*., 2012). The TRUS can sometimes detect tumours that may not have been detected by a DRE, and it may also give clinicians a better idea of PSA density, which can help distinguish between BPH and prostate cancer (Descotes *et al*., 2007).

1.4.4 Biopsy

A biopsy can be ordered as a follow up on a PSA test, DREs and/or or imaging tests. However, there are many factors that should be considered before a biopsy. The patient's medical history, age, ethnicity, heredity, other present diseases, as well as results from any other preceding tests are all factors to consider. A biopsy

is a procedure in which a sample of body tissue is removed and examined under a microscope (Basch *et al*., 2012). A core needle biopsy is the main method used to diagnose prostate cancer. Using transrectal ultrasound to have an image of the prostate gland, a clinician then inserts a thin, hollow needle through the wall of the rectum into the prostate. When the needle is pulled out it removes a small cylinder of prostate tissue. This can be repeated from 8 to18 times (Horwich *et al*., 2010). While biopsies and an analysis of the tumour histology can allow clinicians to appropriately determine the patient's disease and its severity, the biopsy procedure can also lead to adverse events, such as infection, bleeding, and urinary difficulties (Carter *et al*., 2013). There is also a risk of false diagnosis. This can happen if the needle misses the tumour (American Cancer Society, 2012).

Thus, an ideal screening test should be minimally invasive, accurate and conveniently available to the general population. It should also be able to provide an accurate and early diagnosis. The current diagnostic methods for prostate cancer as discussed above are lacking in sensitivity and specificity. They are uncomfortable and maybe potentially harmful not only physically but psychologically and lead to patient apathy towards them. Thus, a molecular marker that is inexpensive, sensitive enough to detect the disease before tumour formation and remains unaltered during disease progression is necessary.

1.5 Biomarkers

The discovery of cancer biomarkers has become an integral part of medicine. A reason why biomarkers have revolutionized the medical field is because they are

effective non-invasive indicator molecules. The National Cancer Institute (NCI) defines biomarkers as any biological molecule that is isolated from body fluids such as urine, blood or tissue that is a sign of a normal or abnormal process in the body or disease condition. They are objective indications of a medical state observed from outside the patient, which can be measured accurately and is reproducible (Strimbu and Tavel, 2010).

Biomarkers can also be used to assess the effectiveness of particular therapies in improving the effects of a disease. By using easily obtained biomarkers to monitor a patient's reaction to a particular drug, it is possible to determine whether treatment is effective for that individual by measuring drug response rates or toxic effects associated with the drug (Kulasingam and Diamandis, 2008).

During any biological process or disease progression, organs secrete biological markers that reflect the occurrence of physiological function (Good et al., 2007). Tumours and cancer cells release RNA, DNA and proteins such as growth factors and cytokines into circulation (De Bock et al., 2009). Thus these released molecules are able to pinpoint the organ they originated from to facilitate the detection of the problem or disease. As a consequence biological fluids such as urine, plasma, serum or cerebrospinal fluid are good sources for mining biological indicators to distinguish a disease state from the non-diseased state (Mayuex., 2004).

Specific clinical biomarkers have the potential to revolutionize the diagnosis and treatment of a variety of medical conditions. The lack of current methods of detection of many diseases at an early stage is a major factor in the prevalence of

diseases. Thus, the objective of biomarker discovery is to identify specific molecular markers susceptible to improve early diagnosis, survey therapeutic outcomes and facilitate the development of novel drug candidates (de Bock *et al.*, 2010). The assumption that the pathology of concern will affect some physiological processes causing changes at the molecular or protein expression levels, is the fundamental approach to biomarker discovery (de Bock *et al.,* 2010).

1.5.1 Types of biomarkers

Cancer biomarkers can be classified into five categories: risk assessment, screening/diagnostic, prognostic, predictive, and pharmacodynamic (figure 1.8). Risk assessment biomarkers use studies on chromosomal aberrations, micronuclei and other changes deemed to represent genomic damage to select for potential disease onset in high risk individuals (Vaini, 2001; Teixeira *et al*., 2105). Prognostic biomarkers predict the natural course of the cancer as well as distinguish the tumour's outcome. Biomarkers also help determine whom to treat, how aggressively to treat, and which candidates will likely respond to a given drug and the most effective dose. Predictive biomarkers evaluate the probable benefit of a particular treatment. Pharmacodynamic biomarkers assess the imminent treatment effects of a drug on a tumour and can possibly determine the proper dosage in the early stages of clinical development of a new anticancer drug (Sawyers, 2008).

Instead of analyzing the tumour cells themselves, the molecular composition of a tumour can be indirectly characterized by analyzing blood samples and searching for variations in serum proteins, thereby improving the precision of screening and curtailing the need for invasive diagnostic procedures. Some difficulties were encountered initially in an attempt to reproduce these cancer associated serum proteins. With advances in our ability to measure quantitatively, collect standardized samples, and resolve the problems of reduced sensitivity in detection, confidence in the results of this approach has risen (Sawyers, 2008). Measurements from biomarkers can be used to adjust empirical results of clinical trials by establishing a relationship between the effects of interventions on molecular/cellular pathways and clinical responses, thereby providing a way for scientists to comprehend mechanistically the differences in clinical response that may be affected by uncontrolled variables (Atkinson *et al*., 2001).

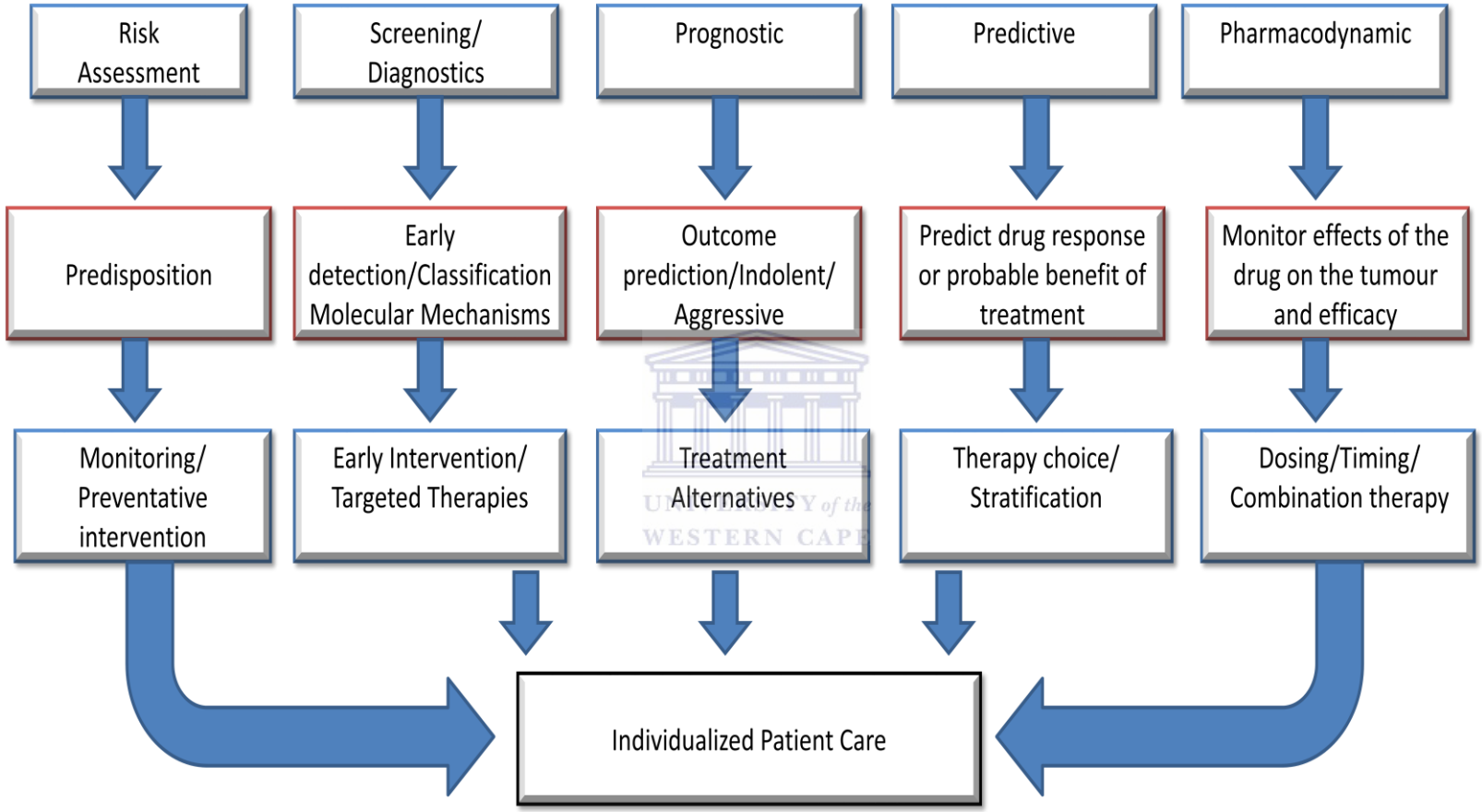


Figure 1.8: Types of biomarkers and their impact on cancer management. (Adapted from Chen *et al*., 2012).

Studies have shown that aberrantly expressed miRNAs are a hallmark of several diseases like cancer (Croce *et al*., 2009). miRNA expression has been shown to be associated with tumour development, progression and response to therapy, suggesting their possible use as diagnostic, prognostic and predictive biomarkers (Iorio and Croce 2012). Because of these characteristics, miRNAs have been emerged as potential as diagnostic, prognostic, and treatment response biomarkers.

1.6 MicroRNAs

miRNAs are naturally occurring endogenous, single stranded RNA molecules. They are 18 -24 nucleotide bases long and are non-protein coding (Munker and Calin, 2013). miRNAs control gene expression by binding to target mRNAs with imperfect complementarity within the 3'-UTR, leading either to their repression of translation or degradation (Cannell *et al*., 2008). They were first discovered in *Caenorhabditis elegans* (*C. elegans*) by Ambros and colleagues in 1993. Since their discovery, they have become the subject of intensive research which has amassed a wealth of information on their biogenesis, function and significance in gene regulation. miRNAs play important roles in a wide range of biological processes including cell proliferation and differentiation, organ development, apoptosis, as well as regulation of several processes related to eukaryotic development (Ardekani and Naeini, 2010). As a consequence, misregulation at any point of these processes owing to abnormal miRNA mutation or expression can result in cancer.

The understanding of the mechanisms of action of miRNAs is still in its infancy. However, to date, work in this area has suggested that miRNAs may regulate gene expression at either the transcriptional or the post-transcriptional level. This may be by either suppressing translation of protein coding genes, or cleaving target mRNAs to induce their degradation (Bartel, 2004).

Currently, there are 2154 mature human miRNAs that have been reported (MiRbase, V21 www.mirbase.org) and they have been predicted to regulate up to 60% of the human protein coding genes.

1.6.1 Biogenesis of MicroRNAs

A larger number of miRNA genes are situated in intergenic regions or in antisense orientation to annotated genes (Lee and Ambros, 2001; Lagos-Quintana *et al*., 2001; Lau *et al*., 2001) this indicates that they are distinct transcription units. There are however, a small number of the human miRNA genes that are located in intronic regions and are transcribed as part of annotated genes (Yang *et al*., 2011).

Biogenesis of miRNAs in humans is a two-step process which involves both nuclear and cytoplasmic cleavage events carried out by Drosha and Dicer (Denli *et al*., 2004; Lee *et al*., 2001). The miRNA genes are transcribed by RNA polymerase into primary polyadenylated miRNA transcripts called pri-miRNAs. Pri-miRNAs consist of one or more hairpin structures each with a stem loop and a terminal loop. They bear a 3' polyadenylated tail and a 5' over hang capped by a 7-methylguanosine. pri-miRNAs can also produce more than one functional

miRNA (Melo and Melo, 2014). Pri-miRNAs are processed into precursor miRNA and subsequently into an miRNA duplex which then releases the mature miRNA (Bartel *et al*., 2004).

The intergenic miRNA is transcribed by RNA polymerases II and III to produce primary miRNA, which as discussed earlier, is a stem looped structure with single stranded RNA overhangs at the 3' end and the 5' end (MacFarlane and Murphy, 2010). Whilst still in the nucleus, a protein complex called the microprocessor comprised of RNase III endonuclease Drosha and the double-stranded RNA-binding protein DiGeorge syndrome critical region gene 8 (DGCR8) (figure 1.9) cleaves the pri-miRNA into a 70 nucleotide hairpin structure called the pre-miRNA (Du and Zamore, 2005). The DGCR8 protein recognizes the stem and the flanking single-stranded RNA and serves as a ruler for Drosha to cut the stem about 11 nucleotides away from the stem–ssRNA junction, releasing the pre-miRNA (Melo and Melo, 2014).

The miRNAs located within introns of protein coding genes are transcribed by RNA polymerase II (Rodriguez *et al*., 2004). Two possible miRNA pathways for the maturation of pri-miRNA have been proposed for intronic miRNAs (Figure 1.9). These processes may occur simultaneously or independently. In one such pathway referred to by some researchers as the non-canonical pathway, some miRNAs known as mirtrons bypass the Drosha step. Instead, introns are excised out of the pri-miRNA by spliceosomal components. The product is a pre-miRNA that continues to the cytoplasm for maturation (Kim *et al*., 2007). The alternative

pathway, called the canonical, is the same as that of the intergenic miRNAs, involving cleavage with Drosha (Melo and Melo, 2014).

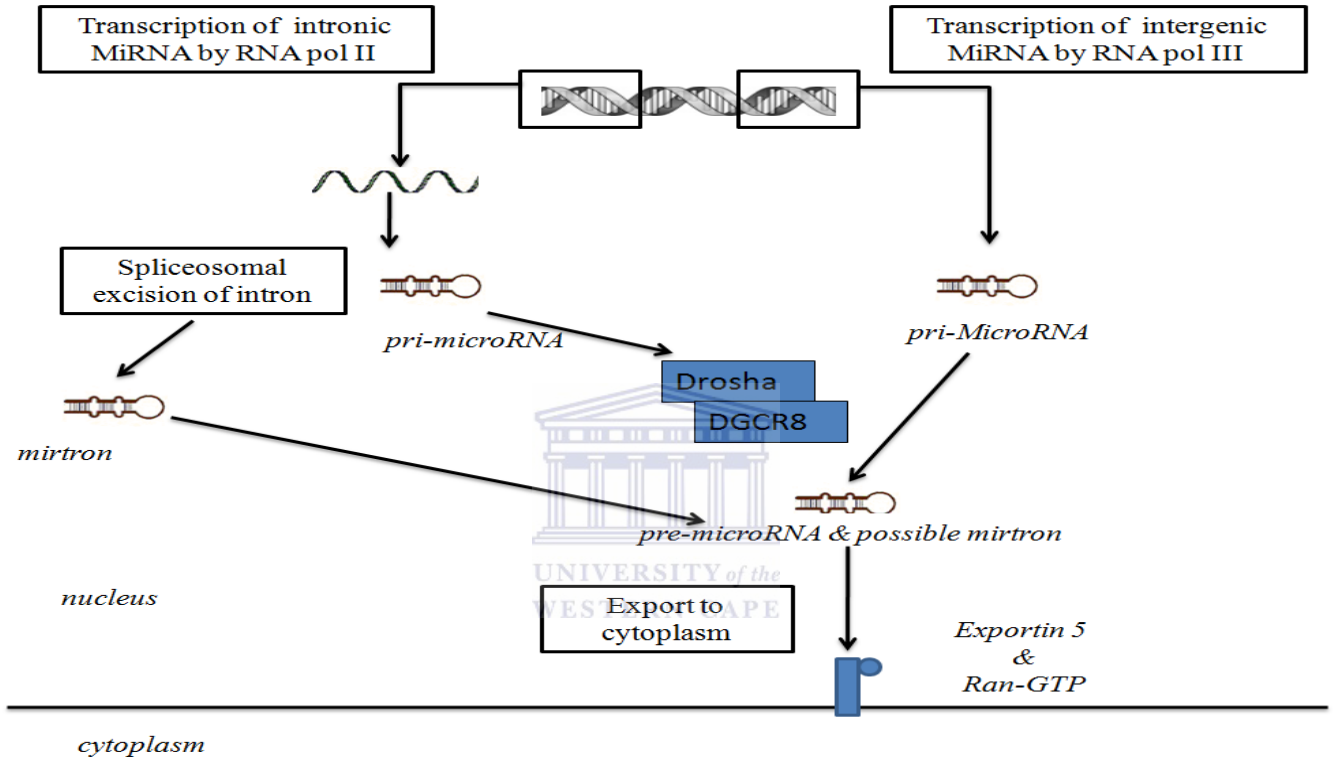


Figure 1.9: Nuclear section of miRNA biogenesis.
Intergenic miRNAs are transcribed by RNA polymerase III producing a pri-miRNA molecule, which is processed into a precursor miRNA (pre-miRNA) by the microprocessor complex comprised of DGCR8 and Drosha. Pre-miRNAs are exported to the cytoplasm by Exportin 5 and Ran-GTP which are nucleocytoplasmic transporters. Intronic miRNA are transcribed by RNA poly II. The miRNA sequence is excised from the pri-mRNA by spliceosomal components to liberate a mirtron which is exported to the cytoplasm. (Adapted from MacFarlane and Murphy, 2010).

The export of the pre-miRNA from the nucleus to the cytoplasm is facilitated by the nucleocytoplasmic transporter factor Exportin-5 which is a RanGTP-dependent dsRNA binding protein (Bohnsack et al., 2004). At this point, the pre-miRNA is a stem looped structure about 60-70 nucleotides long. In the cytoplasm, as seen in figure 1.10, an RNase III enzyme Dicer, in a complex with its binding partner the transactivator RNA-binding protein (TRBP), binds to the pre-miRNA on the second nucleotide at the 3' overhang. Dicer then cleaves off the terminal base pairs and the loop of the pre-miRNA. This results in a double stranded RNA structure called the miRNA:miRNA* duplex with 3' overhangs. This structure contains the mature miRNA strand (22 nucleotides in length) and its complementary strand denoted by an asterisk (Li *et al*., 2007). The mature miRNA strand is then loaded onto miRISC, which is a miRNA associated RNA-induced silencing complex. miRISC contains an Argonaute protein (Ago 2) which catalyzes the cleavage of mRNAs (figure 1.10). Thus it is also a protein responsible for translational repression (MacFarlane and Murphy 2010). The complementary strand is degraded.

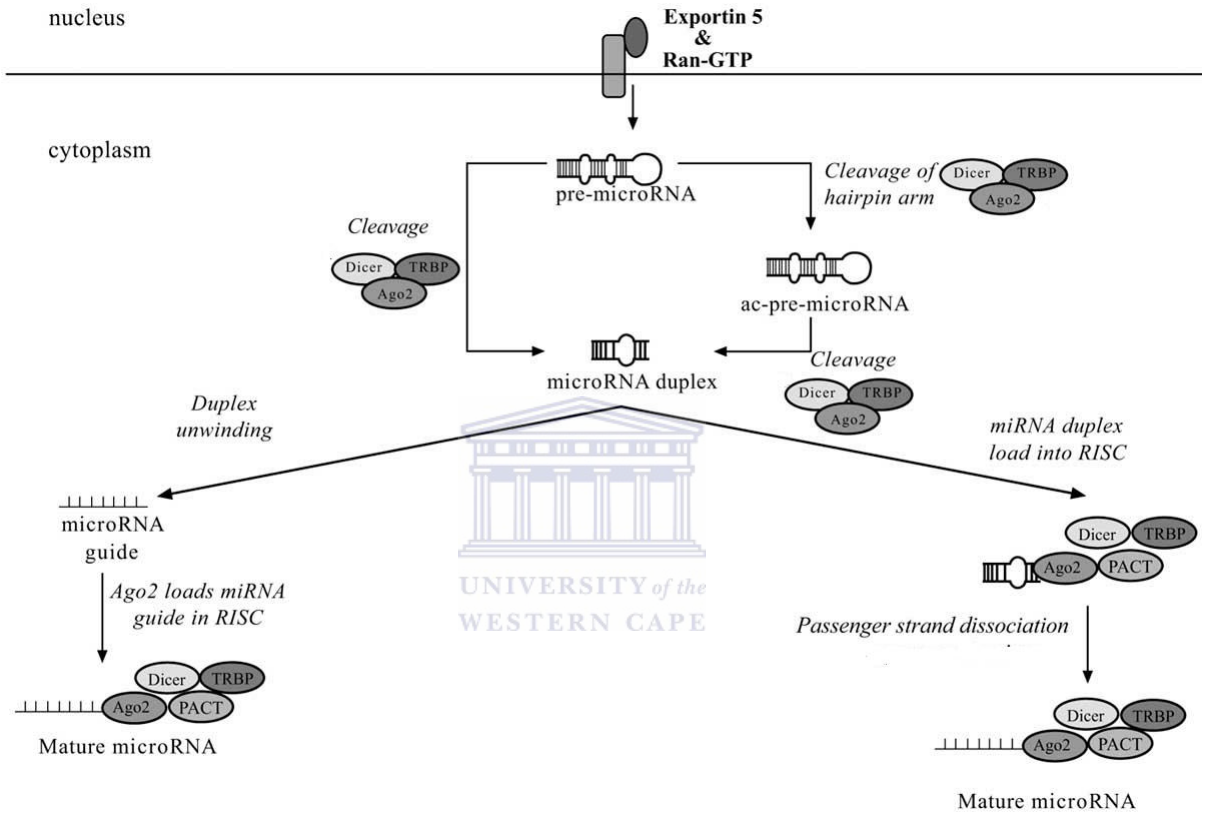


Figure 1.10. Cytoplasmic component of microRNA biogenesis.
Pre-miRNA is cleaved by Dicer to generate an miRNA duplex. The miRNA duplex liberates the mature miRNA to assemble into a RISC loading complex comprised of Ago2, TRBP, PACT and Dicer. The mechanism of mature miRNA release is unclear. (Adapted from MacFarlane and Murphy, 2010).

The mechanism of the miRNA and RISC assembly is not well understood in humans. There are many hypotheses with regards to the unwinding of the duplex. There is evidence however, that the process is ATP-independent. Gregory *et al*., 2005, hypothesised that Dicer cleaves the passenger strand and thus initiates the unwinding of the duplex, releasing the mature single strand that is captured by Ago2. Another hypothesis by Maniataki *et al*., 2005, is that conformational changes in miRISC during assembly could also cause the duplex to unwind simultaneously. The mature miRNA incorporated into the RISC guides the complex to target mRNAs with complementary sequences to inhibit their translation via mRNA cleavage, mRNA destabilization or translational repression (Bartel, 2004).

1.6.2 Regulation of miRNA biogenesis.

As discussed earlier, miRNAs play important roles in a wide range of biological processes such as cell differentiation, organ development, and apoptosis. Therefore, stringent control of their levels is critical to maintaining normal cellular functions and any deregulation of miRNA expression is often associated with human diseases, such as cancer (Jiang *et al*., 2009). As such, it is highly important to have the biogenesis of miRNAs subject to regulation at various levels, from transcription, processing, sequence identity, and binding to target mRNAs.

Transcription is a major point of regulation in miRNA biogenesis. Many miRNA gene promoters have characteristics that are similar to protein-coding gene promoters. Therefore, the same mechanisms of gene expression control such as use of silencers, specificity factors and repressors can be employed (Ozsolak *et al*., 2008; Corcoran *et al*., 2009). miRNA genes can be transcribed by either RNA polymerase II or III. A wide range of regulatory options can be facilitated as each enzyme is regulated differently and recognizes specific promoters and terminators (Melo and Melo, 2014).

Regulation can also occur during the Drosha processing pathway. Drosha and DGCR8 levels in the cell are tightly controlled and as such play an important role in the regulation of pri-miRNA processing (Davis *et al*., 2008). The protein DGCR8 has a stabilizing effect on Drosha by interacting with its middle domain, and Drosha can control the levels of DGCR8 by cleaving hairpins in mRNA coding for DGCR8 thus leading to its degradation (Triboulet *et al*., 2009). Therefore, a well controlled equilibrium needs to exist between the amounts of Drosha and DGCR8 to maintain the processing of pri-miRNAs.

1.6.3 Function of miRNAs

miRNAs play a vital role in regulating numerous metabolic and cellular pathways, notably those controlling cell proliferation, differentiation and survival (Bushati and Cohen, 2007). For these purposes, an miRNA is complementary to one or more mRNAs. In animals the complementary site is usually in the 3'-UTR (Cannell *et al*., 2008). Perfect or near perfect complementarity promotes cleavage

of the mRNA. However, nucleotides 2-7 of the miRNA have to be perfectly complementary (Lagos-Quintana *et al*., 2001). Binding of miRNA to the mRNA can inhibit protein translation or accelerate the process of deadenylation causing the mRNA to be degraded (Melo and Melo, 2014).

Dysregulation of miRNA expression has been demonstrated in most tumours examined (Gong *et al*., 2005). However, the specific classification of miRNA as oncogenes or tumour suppressors can be difficult because of the intricate expression patterns of miRNAs. These expression patterns differ for specific tissues and differentiation states and this poses two difficulties in classification (Bartel *et al*., 2004; Jiang *et al*., 2009). Firstly, it is not always clear if altered miRNA patterns are the direct cause of the cancer or rather an indirect effect of changes in cellular phenotype. Secondly, a single miRNA can regulate multiple targets (MacFarlane and Murphy, 2013). This, coupled with tissue specific expression could, implicate a single miRNA as a tumour suppressor in one context and an oncogene in another.

1.6.4 miRNA regulation in cancer

Since the discovery of miRNAs in 1993, their regulation has been involved in a large variety of physiological processes. miRNAs were first reported to be dysregulated in cancer in 2002, when Calin and co-workers described a deletion at 13q14 in chronic lymphocytic leukaemia. The miR-15/miR-16 clusters are located on this locus and are not only down-regulated in these tumours but also

regulate B-cell lymphoma 2 expression. Early studies also suggested that there was a predominant down regulation of miRNAs (Lu *et al.*, 2005) and DICER expression down regulation in cancer supporting this hypothesis (Karube *et al.*, 2005). However, further studies revealed both up and down regulation of miRNA in all solid tumours and leukaemias (Calin *et al*., 2002). However, some miRNAs seem to have a tissue-specific function and are only expressed in specific cancers; others are universally over or under-expressed in cancer. For example, miR-21 has been shown to be up-regulated in almost all solid tumours, including prostate, breast, lung, pancreas, stomach, and colon tumours (Chan *et al.*, 2005; Karube *et al.*, 2005).

1.6.5 miRNAs as tumour suppressors and oncogenes

It is well documented that up-regulation or down-regulation of miRNAs occurs in various human cancers (Zhang *et al*., 2007, Vaz *et al*., 2013). Over-expressed miRNAs may function as both oncogenes (through down-regulation of tumour-suppressor genes) and/ or regulators of cellular processes such as cell differentiation or apoptosis (He *et al*., 2005, 2007).

Let-7 miRNAs were one of the first miRNAs discovered in *Caenorhabditis elegans* (Ardekani and Naeini, 2010). Let-7s are highly conserved among invertebrates and vertebrates, including humans who have twelve let-7 genes encoding nine miRNAs (Zhang et al., 2007; Ardekani and Naeini, 2010). Several let-7 genes have been mapped to regions within the human genome that are

frequently altered or deleted in various cancers (Calin *et al*., 2004). This discovery has implicated let-7 as a tumour suppressors.

Two well-defined let-7 targets are the oncogenes, Ras and high mobility group AT-hook 2 (HMGA2). Ras is a signal transducing GTPase that delivers signals from cell surface receptors to functional intracellular pathways, thus affecting cell proliferation, growth, cytoskeleton organization, cell movements and survival (Pylayeva-Gupta *et al*., 2011). As a result, active Ras mutants (H-Ras, K-Ras and N-Ras) are found within a variety of human cancers including pancreatic, colon, thyroid and lung carcinomas. Let-7s regulate the expression of Ras, and its mutants via 3' UTR binding which inhibits translation. Thus, when functioning as a tumour suppressor, let-7 mediates the suppression of Ras and its cellular processes (He *et al*., 2007; Pylayeva-Gupta *et al*., 2011).

HMG2A is a non-histone architectural transcription factor that alters DNA conformation, leading to direct transcriptional activation of a variety of genes that influence cell growth, differentiation, proliferation and survival (Wantanbe *et al*., 2009). HMG2A is undetectable in normal adult tissue but is highly expressed in embryonic tissues, lung cancer and uterine leiomyomas (Wantanbe *et al*., 2009; Zhang *et al*., 2007). Studies reveal that the let-7s regulate HMG2A by destabilizing its mRNA through 3'-UTR binding (Zhang *et al*., 2007).

The let-7 family of miRNAs regulate numerous genes not all of which are as well defined as HMG2A and Ras. Recent data indicates that let-7s play a much larger role in controlling cell proliferation than initially thought as they have been

shown to functionally inhibit numerous cell cycle regulators including c-myc, CDC25A, CDK6 and cyclin D2 (Worringer *et al*., 2014; Vaz *et al*., 2013).

1.6.6 miRNAs as molecular markers

The frequent deregulation of miRNAs in cancer makes them attractive new markers for cancer detection and monitoring. miRNAs can either serve as diagnostic, prognostic or predictive markers; they can also monitor therapy success. Studies have suggested that miRNAs can help to differentiate cancer from normal tissue, discriminate different tumours subtypes, characterize poorly differentiated tumours and identify tumours of unknown origin (Lu *et al.*, 2005). The stability of miRNAs, provides an advantage in using them over other markers. Studies have shown that miRNAs can be stably expressed in degraded total RNA samples from human tissues (Karube *et al.*, 2005), and they remain detectable in formalin-fixed, paraffin-embedded tissue (Arroyo *et al.*, 2011; Yaman *et al.*, 2008, 2011).

miRNA measurements in body fluids, such as blood or urine, are crucial for diagnostic purposes because these are fluids that are more accessible as compared to tissue or organs (Chan *et al.*, 2005). miRNAs are present in body fluids as free miRNAs in exosomes (Michael *et al.*, 2010), associated with AGO2 proteins (Arroyo *et al.*, 2011) or in circulating tumour cells (Zhang *et al.*, 2011). Studies by Yaman *et al.*, 2011 and Zhang *et al.*, 2011 indicated that miR-141 and miR21 circulating miRNAs, may serve as non-invasive markers in prostate cancer.

Lawrie *et al*., 2008, first reported the existence of circulating miRNAs in serum and their possibility to be developed as cancer markers. The mechanism of release and the roles of circulating miRNAs are still largely unclear. Recently, Skog et al., 2008 showed that the interaction between cells via mRNA and miRNA can be accomplished by microvesicle transfer, thus circulating miRNAs are proposed to be released from tumour cells in microvesicles to ensure communication with recipient cells in the surrounding microenvironment. Because circulating miRNAs might manipulate target cells, cell-free miRNAs might be not only serum markers for cancer and disease progression but also functionally relevant and therefore potential targets for novel therapy approaches (Brase *et al*., 2010).

MiR-21 is one of the first miRNAs to be described as an oncomir. As most of the targets of miR-21 are tumour suppressors, miR-21 is associated with a wide variety of cancers including that of the breast, ovaries, cervix, colon, brain and prostate. Studies by Brase *et al*., 2011; Li *et al*., 2012 and Jackson *et al*., 2014 showed that miR-21 expression in prostate cancer tissue and serum samples is significantly associated with the pathological stage of prostate cancer as well as lymph node metastasis.

Brase *et al*., 2011 also examined miR-141 as a miRNA correlated with tumour progression in cancer. The study identified it as one of the miRNAs present at high levels in patients with malignant tumours when compared to PSA. In the same study, miR-141 was also shown to distinguish between patients with metastatic prostate cancer and healthy controls. However, both miR-21 and miR-

141 are implicated in other cancers apart from prostate cancer and thus may not be suitable diagnostic markers for the disease but rather have prognostic value.

1.7 Bioinformatics as a tool for the detection of novel biomarkers

Over the past years, major advances have been accomplished in the field of molecular biology and these have been linked with advances in high throughput technologies such as genomics, transcriptomics and proteomics (Emmett *et al*., 2014) Furthermore, these technologies have brought forward an explosive amount of biological information which has led to the need for computerised databases to store, organise, and analyse the data (Martone *et al*., 2004; Benson *et al*., 2012). For this reason, the field of bioinformatics, or systems biology, which is the merging of the computational and biological science disciplines, has been an important tool for the organisation and analysis of the vast amount of biological data (Zhang *et al*., 2011). The main aim of bioinformatics is to find key biological information hidden amongst a mass of raw data to identify important trends and patterns which would eventually lead to novel biomarker discovery for both diagnostic and therapeutic purposes (Raza, 2012). Additionally, bioinformatics allows for the *in silico* simulations of complex disease physiologies, such as interactions between components, on their molecular level (Berman *et al*., 2013). Bioinformatics has presented ways in which data mining approaches can be used to filter valuable targets such as miRNA, genes, or proteins for the discovery of possible novel biomarkers for diseases (Sommer *et al*, 2010).

1.9 Role of bioinformatics in miRNA research

Studies have shown that aberrantly expressed miRNAs are a hallmark of several diseases like cancer (Croce *et al*., 2009). miRNA expression has been shown to be associated with tumour development, progression and response to therapy, suggesting their possible use as diagnostic, prognostic and predictive biomarkers (Iorio and Croce 2012). Identifying miRNAs, their target genes, and their respective regulatory functions are important for understanding normal biological processes as well as understanding their various roles in disease development (Zhang and Verbeek, 2010; Liu *et al*., 2012; Fujiwara and Yada, 2013). Bioinformatics facilitates experimental validation of miRNAs and their target genes by producing statistically significant hypotheses from biological data that has been stored in databases, based on other biological experimental data (Liu *et al*., 2012). Potential target identification is based on the software's algorithm. There are several miRNA-target prediction software tools publicly available (Fujiwara and Yada, 2013) which include TargetScan, and miRDB (Wong and Wang 2015). Possible targets are predicted based on the software's prediction algorithms. The target prediction software as well as the other platforms used in this study will be discussed further in the next chapter.

1.10. Study rationale

The principal involvement of miRNAs in the aetiology and progression of many common diseases indicates these molecules are significant markers with potential use as diagnostic, prognostic and therapeutic tools. The discovery that miRNAs

are detectable and quantifiable in the circulation of diseased persons adds further validity to their potential as biomarkers of disease, both benign and malignant. The investigation of existing miRNA molecules coupled with the identification of novel miRNAs, and elucidation of their downstream targets, will provide a better understanding of their functional effects and thus provide greater insight into the complex and poorly understood mechanisms underlying diseases such as cancer.

In PCa, the roles of miRNAs have become clear due to the understanding of the interactions between miRNAs and their targets and the resulting impact on prostate carcinogenesis. Thus the purpose of this study was to identify specific miRNAs as potential early diagnostic biomarkers in prostate cancer using a combination of *in silico* methods and molecular methods. Specific study aims are outlined as follows:

1.) The *in silico* identification of specific miRNAs and their targets involved in prostate cancer progression, as well as categorization of gene networks these targets are involved in, and the implications of the identified miRNAs in cancer causing pathways.

2.) Analyse expression profiles of identified miRNAs in prostate cancer cell lines as well as a panel of cancer cell lines including breast, liver, colon, lung and oral carcinomas using real time PCR (qRT-PCR).

3.) Analysis of identified prostate miRNAs as potential early diagnostic biomarkers by investigating whether the panel of miRNA identified are PCa specific or generalised to several malignancies studied via the cell lines used in this work.

Chapter 2

Identification of miRNAs as biomarkers for the detection of Prostate cancer (PCa) using an *in silico* approach

2.1 Introduction

The sequencing of the genomes of many organisms as well as the human genome brought about the availability of a large amount of genomic data. In an attempt to make use of this vast wealth of data from the genome sequencing projects, new fields in molecular biology such as functional genomics, proteomics and metagenomics were established (Hancock *et al*., 2013).

These fields attempt to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products accomplished by the study of patterns of gene expression under various conditions (Raza, 2012). Advancements in technologies involved in these fields such as DNA and protein microarrays, have resulted in an explosion of large scale experimental and literature data. Thes data are stored in various repositories known as databases (Raza, 2012). These databases perform a myriad of functions such as the analysis of gene and protein expression and regulation, gene annotations, comparison of genetic and genomic data, as well as the simulation and modelling of DNA, RNA, and protein structures as well as molecular interactions in structural biology (Zhang *et al*., 2011). As a result, the past 15 years have seen an exponential increase in these databases (Raza, 2012).

These databases include public repositories such as the GenBank, a gene database hosted by NCBI (National Centre for Biotechnology Information) (Benson *et al*., 2012) and the Protein DataBank (PDB), a protein database hosted by the Worldwide Protein Data Bank (Berman *et al*., 2003). There are also a number of private databases, many of which are hosted and used by biotech companies or research groups involved in gene mapping projects (Benson *et al*., 2012). In this study, a number of databases were used for various objectives including identifying the target genes of miRNAs implicated in prostate cancer and analysing the interaction of proteins encoded by these genes.

2.2 Data mining

The process of retrieving interesting patterns of information from the aforementioned databases is referred to as data mining. It can also be categorised as the extraction of information from different sources of published literature (Yang *et al*., 2009). Data mining has also been used to identify disease associated entities and to understand their roles in disease onset as well as progression. Because of the unprecedented growth of genomic data, data mining has become a field of interest as almost any new discovery or development about a gene, its pathways or protein it codes for is recorded in literature and consequently updated in databases. Thus the goal of data mining is to filter that knowledge and to present the resulting information to a user in a concise, understandable format (Faro *et al*., 2012).

The use of computational tools in data mining to decipher biological data has resulted in new methods for biomarker and disease discovery, as well as disease pathway elucidation. An *in-silico* approach to target discovery for early diagnosis, understanding of disease and development of therapeutics can save considerable wet bench time. This approach can reveal important evidence for evolutionary and functional relationships between genes and proteins (Krallinger and Valencia 2005).

2.3 Pathway Analyses using mirPath tool in DIANA-TarBase version 7

The DIANA-miRPath is a tool that performs miRNA pathway analysis, providing accurate statistics. It is available at the DIANA-TarBase database (DIANA miRPath version 7.0) accessible at http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index. The tool is capable of combining results with merging and meta-analysis algorithms, performing hierarchical clustering of miRNAs and pathways based on their interaction levels, as well as being able to generate sophisticated visualization of results, such as dendrograms or miRNA versus pathway heat maps (Paraskevopoulou *et al.*, 2013).

DIANA-miRPath hosts numerous novel features, extensions and optimizations, including: (1) the use of predicted interactions derived from DIANA-microT-CDS (Vlachos *et al.*, 2014) and TargetScan 6.2 (Hsu *et al.*, 2014); (2) a significant extension to the annotation database, enabling the identification of

miRNAs controlling molecular pathways as well as the performance of miRNA function annotation using gene ontology (GO) terms (Xiao *et al*., 2009; Paraskevopoulou *et al*., 2013); (3) a new Reverse Search Module with unprecedented flexibility that can assist in (re)-discovering miRNAs with not yet identified functions; and (4) support for seven model species: *H. sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Gallus gallus* and *Danio rerio*.

The gene and miRNA annotations in DIANA-mirPath are derived from Ensembl (Paraskevopoulou *et al*., 2013) and miRBase (*Vlachos et al*., 2014), respectively. The miRNA: gene interactions are derived from the *in silico* miRNA target prediction algorithms: DIANA-microT-CDS and TargetScan 6.2, the latter operates in both Context+ and Conservation modes (Paraskevopoulou *et al*., 2013). DIANA-microT-CDS is a highly accurate target prediction algorithm with target prediction in 3′ UTR and CDS mRNA regions.

2.4 Target gene prediction

Many studies have shown that miRNAs play an essential role in gene regulatory networks (Section 1.5.3) by controlling the expression of genes involved in important biological processes (Bartel, 2009). Thousands of miRNA genes have been identified, but the functions of most of these miRNAs remain unknown due to the lack of experimental and computational approaches to predict their exact target mRNAs (Agarwal *et al*., 2015).

miRNA functional characterization is currently a very active research field in biology, and there has been a rapid accumulation of miRNA knowledge in the past few years. Although a large quantity of existing data is certainly very helpful to guide future studies, it can at the same time make it challenging for miRNA researchers to quickly retrieve information relevant to their studies. Because of this, several miRNA databases have been established to systematically organize miRNA data (Grimson *et al*., 2007). The most prominent one is miRBase, which was set up to provide official nomenclatures to the miRNA research community (Griffiths-Jones *et al*., 2006). Although miRBase is a valuable source for providing standard miRNA nomenclature, it contains limited information on miRNA functional annotation. Thus, besides miRBase, a few other databases have been developed to focus more on miRNA function and consequently target prediction. Some examples of these databases include TargetScanHuman, miRDB, and PicTar. These databases host miRNA targets predicted by different computational algorithms (Krek *et al*., 2005; Grimson *et al*., 2007). However, common features of miRNA target prediction tools include, seed match conservation, free energy and site accessibility. This section introduces two databases for miRNA gene target prediction in humans, TargetScanHuman and miRDB which were used in this study to improve the robustness of the predictions.

2.5. TargetScanHuman

Early studies of target recognition revealed that near-perfect complementarities at the 5' end of a miRNA, the so-called "seed regions" at positions 2 to 8, are

primary determinants of target specificity (Bartel *et al*., 2004). These sites are called the canonical binding sites and binding happens through Watson and Crick base pairing (Agarwal *et al*., 2015). However, studies have also shown that binding these regions may not always be sufficient for gene repression (Grimson *et al*., 2007).

TargetScanHuman is a web server that predicts biological targets of miRNAs by searching for the presence of conserved 8-mer, 7-mer and 6-mer sites that match the seed region of the 3' UTR of a mRNA (Agarwal *et al*., 2015). It can be launched at http://www.targetscan.org/. TragetScanHuman not only predicts mRNAs that have the aforementioned conserved sites, it also considers the flanking sites of the seed region. A study by Grimson *et al*., 2007 found that effective binding sites reside within an AU-rich region. Thus, TargetScanHuman takes advantage of this in its prediction algorithm (Agarwal *et al*., 2015).

Additionally, an accurate ranking of the predicted targets for each miRNA is provided (Friedman *et al*., 2009). This ranking is based on either the probability of evolutionarily conserved targeting (Agarwal *et al*., 2015) or the predicted efficacy of repression calculated using cumulative weighted context ++ scores of the sites (Agarwal *et al*., 2015). Predictions are also ranked by their probability of conserved targeting (Friedman *et al*., 2009). TargetScanHuman uses the miRNA ID information as a query to extract target gene information from all available experiments publicly available.

2.6 miRDB

miRDB is a freely accessible online tool for miRNA target prediction and functional annotations (Wong and Wang, 2015). It can be launched from http://mirdb.org. A feature of this platform is the MirTarget algorithm, which was developed by analyzing thousands of miRNA target interactions from high-throughput sequencing experiments. Unlike TargetScanHuman, seed conservation is a consideration and not priority in miRDB. Seed match considerations are only in the 7-mer and 8-mer regions. miRDB is also different from existing miRNA target predicting databases. Firstly, its database design strategy is centred on mature miRNAs. If a miRNA has three pre-cursors in a genome, there are three different databases for it (Wang, 2008). Thus, there is no centralized place to present the miRNA functional annotations (Wang, 2008; Peterson *et al*., 2014). As mature miRNAs are the carriers of miRNA function, miRDB is designed to focus on them only to avoid database redundancy (Wang, 2008; Peterson *et al*., 2014; Wong and Wang, 2015). Secondly, miRDB has a wiki editing interface for community-provided miRNA annotations (Wang, 2008; Wong and Wang, 2015). With the rapid progress on miRNA functional studies, it is challenging for any single team to keep track of all the latest developments in the field (Wong and Wang, 2015). Thus, miRDB allows miRNA researchers to provide miRNA functional annotations and actively interact with each other (Wang, 2008).

2.7 Functional characterization of genes via DAVID

DAVID (**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery) is a publicly available high-throughput annotation tool that maps a large number of genes to a list of associated Gene Ontology terms and statistically highlights genes that are highly enriched for those terms (Huang *et al*., 2009). This increases the potential to identify biological processes most pertinent to the biological phenomena under study. DAVID is available at https://david.ncifcrf.gov/ version 6.7 (Jiao *et al*., 2012).

Gene ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species (Rancaglia *et al*., 2013). The gene ontology project specifically aims to: 1) maintain and develop its controlled vocabulary of gene and gene product attributes; 2) annotate genes and gene products, and assimilate and disseminate annotation data; and 3) provide tools for easy access to all aspects of the data provided by the project. It also enables functional interpretation of experimental data using the GO, for example via enrichment analysis (Gene Ontology Consortium, 2004).

The gene ontology covers three domains: 1) Cellular component, the parts of a cell or its extracellular environment. 2) Molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis and 3) Biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units such as cells, tissues, organs, and organisms (Gene Ontology Consortium, 2004; Rancaglia *et al*., 2013).

There are many functions performed by the DAVID platform including functional annotation, gene functional classification and gene ID conversion. For the purpose of this study, the functional annotation tool was employed.

One tool DAVID has is the Functional Annotation Clustering tool which is a newly added feature to the DAVID Functional Annotation Tool (Huang *et al*., 2009). This function uses a novel algorithm to measure relationships among the annotation terms based on the degree of their co-association genes to group the similar, redundant and heterogeneous annotation contents from the same or different resources into annotation groups (Huang *et al*., 2009). This reduces the burden of associating similar redundant terms and makes the biological interpretation more focused on a group level. The tool also provides a look at the internal relationship among the clustered terms. The clustered format is able to give a more insightful view about the relationships of annotations (Dennis *et al*., 2003).

2.8 Gene/protein interaction analysis via STRING

Complete knowledge of all direct and indirect interactions between proteins in a given cell would represent an important milestone towards a comprehensive description of cellular mechanisms and functions. Currently, to achieve this goal is elusive. However, considerable progress has been made; particularly for certain model organisms and functional systems (Madu and Lin, 2010). At present, protein interactions and associations are annotated at various levels of detail via online resources, ranging from raw data repositories to highly

formalized pathway databases (Franceschini *et al*., 2013). For many applications, a global view of all the available interaction data is desirable, including computational predictions.

One such online database that predicts protein-protein/gene interactions is STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) http://string-db.org/.

STRING aims to provide a comprehensive, yet quality controlled collection of protein-protein associations for a large number of organisms (Snel *et al*., 2003). The associations are derived from high throughput experimental data, from the mining of databases and literature, and predictions based on genomic context analysis (Mering *et al*., 2005). STRING integrates and ranks these associations by benchmarking them against a common reference set, and presents evidence in a consistent and intuitive web interface. Importantly, the associations are extended beyond the organism in which they were originally described, by automatic transfer to orthologous protein pairs in other organisms, where applicable. STRING currently holds 730 000 proteins in 180 fully sequenced organisms (Mering *et al*., 2005). STRING has three unique features for protein interaction prediction (1) it provides uniquely comprehensive coverage, with over 1000 organisms, 5 million proteins and more than 200 million interactions stored; (2) it is one of very few sites to hold experimental, predicted and inferred interactions, together with interactions obtained through text mining; and (3) it includes a wealth of accessory information, such as protein domains and protein structures, improving its day-to-day value for users (Franceschini *et al*., 2013).

2.9 Analysis of tissue-specific gene expression profiles via TiGER and GeneHub-GEPIS

2.9.1 TiGER

TiGER (Tissue-specific Gene Expression and Regulation) available at http://bioinfo.wilmer.jhu.edu/tiger/ is a database that contains tissue-specific gene expression profiles or expressed sequence tag (EST) data, cis-regulatory module (CRM) data, and tissue specific transcription factor interaction data in 30 human tissues for each gene contained within the database (Liu *et al*., 2008). At present the database contains expression profiles for 19,526 UniGene genes, combinatorial regulations for 7,341 transcription factor pairs and 6,232 putative CRMs for 2,130 RefSeq genes (Liu *et al*., 2008).

The gene expression pattern for each UniGene is calculated based on the NCBI EST database. TiGER has identified and catalogued 7261 tissue-specific genes for 30 human tissues based on their expression enrichment and statistical significance. Thus, on average, each tissue expresses approximately 290 tissue-specific genes (Yu *et al*., 2007).

In addition to EST data, TiGER also identifies transcription factors (TFs) based on patterns of co-occurrence of pairs of DNA binding sites (Yu *et al*., 2007). TiGER predicts 9060 tissue-specific TF interactions, around 300 for each tissue (Yu *et al*., 2005). To evaluate these results, the database uses known interactions as positive controls due to the scarcity of tissue-specific interactions.

Cis-regulatory modules (CRMs) are a stretch of DNA usually about 100-1000 base pairs in length where several transcription factors can bind (Davidson, 2004). They are the central cis-elements that control gene expression (Istrail and Davidson, 2005). TiGER calculates the interaction strength between two transcription factor (TF) binding sites and then derives an empirical "potential energy" for each TF binding site (Yu *et al*., 2007). This results in energy profiles for the promoter sequences of tissue-specific genes. An energy level less than -1 indicates the existence of a TF module (Liu *et al*., 2008).

This development of computational methods for tissue-specific combinational gene regulation, based on transcription factor binding sites, CRMs and ESTs enables the platform to perform a large-scale analysis of tissue-specific gene regulation in human tissues.

2.9.2 GeneHub-GEPIS

GeneHub-GEPIS is a web application that performs digital expression analysis in human and mouse tissues based on an integrated gene database (Zhang *et al*., 2007). It is available at http://share.gene.com/Research/genentech/genehub-gepis/genehub-gepis-search.html.

The platform calculates the normalized gene expression levels across a large panel of normal and tumour tissues, thus providing rapid expression profiling for a given gene. The Digital Expression Unit (DEU) is used as a measure of

expression profiles in comparison to normal tissue and cancer tissue. For a given gene in each tissue category, it is defined as the number of matching expressed sequence tag (EST) clones from a normalized library size of 1 million (Zhang *et al*., 2007). The backend GeneHub component of the application contains pre-defined gene structures derived from mRNA transcript sequences from major databases and includes extensive cross-references for commonly used gene identifiers (Zhang *et al*., 2004: 2007). ESTs are then linked to genes based on their precise genomic locations as determined by the Genomic Mapping and Alignment Program (GMAP). In addition, the gene-centric design makes it possible to add several important features, including text-searching capabilities, the ability to accept diverse input values, expression analysis for microRNAs, basic gene annotation, batch analysis, and linkage between mouse and human genes (Zhang *et al*., 2007).

2.10 Previous work

A list of 13 miRNAs implicated in prostate cancer was generated in a parallel study, (Khan, 2015), using publicly available databases, MiRBase (http://www.mirbase.org/search.html) (Grifiths-Jones *et al*., 2006) and the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) (Barrett *et al*.,2013). Of the 13 miRNAs, 8 were carried forward in the parallel study and the remaining 5 were used in this study.

2.11 Aims and Objective

This chapter aimed to identify putative miRNA biomarkers associated with prostate cancer as well as explore some predicted interactions and pathways these miRNAs are involved in. This is in an effort to identify putative genes as prostate cancer biomarkers in conjunction with the targeting miRNAs. The work outlined in this chapter was undertaken using *in silico* methods.

The specific study objectives were:

i) Predict and analyse participating pathways for the 5 identified miRNAs using DIANA-Tarbase.

ii) To identify the genes targeted by the 5 miRNAs identified from the previous study using TagetScanHuman and miRDB.

iii) Functionally annotate the miRNA targeted genes using DAVID and furthermore generating gene/protein interactions networks using STRING.

iv) To perform an *in silico* expression analysis of the miRNA targeted gene using TiGER and GeneHub-GEPIS.

2.12 Methodology

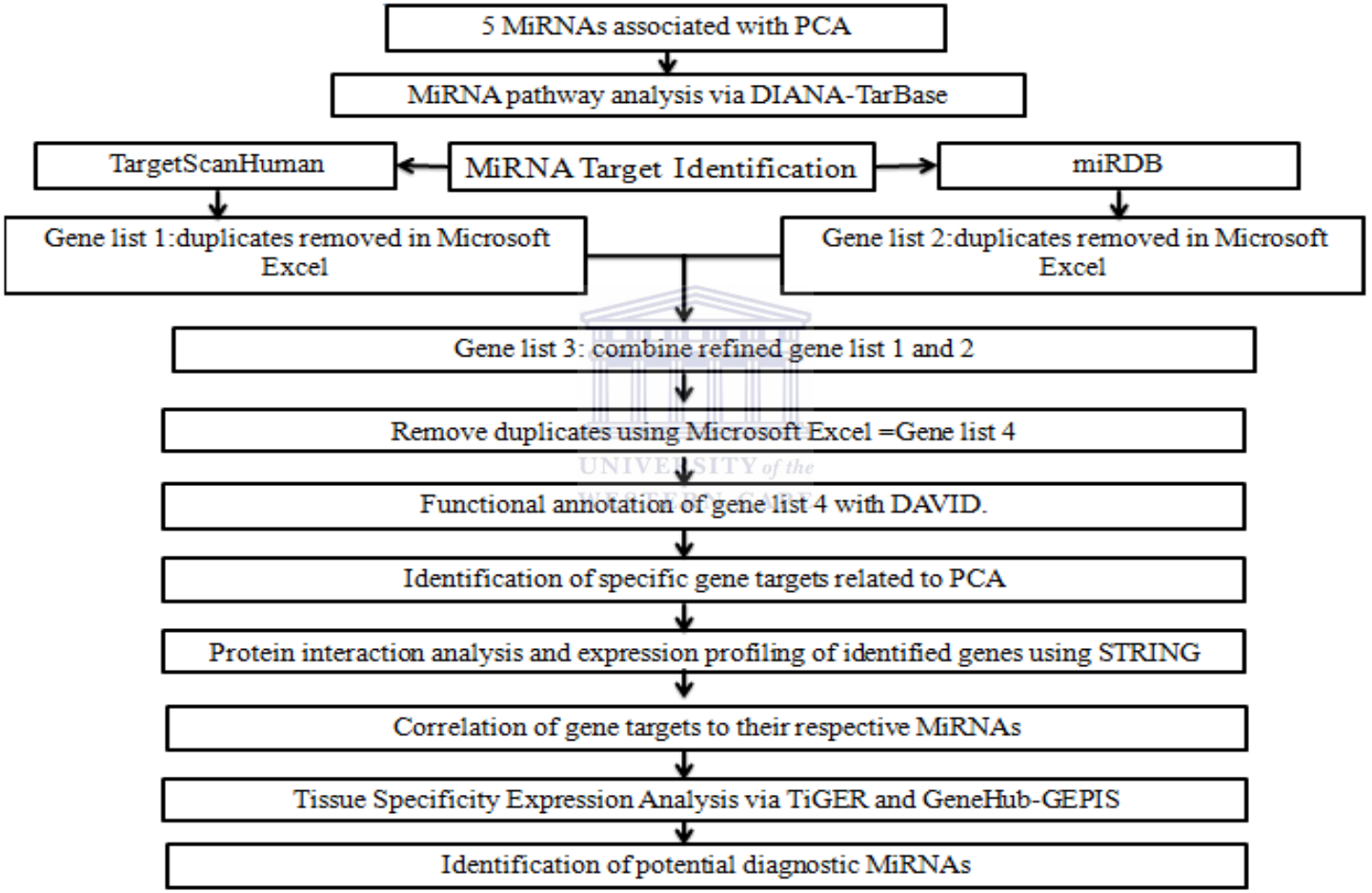


Figure 2.1: Outline of the *in silico* methodology for prostate cancer miRNA biomarker discovery.

2.12.1 Pathway Analyses using mirPath Tool in DIANA-TarBase Version 7

Pathway analysis for this section of the study was achieved using the mirPath software tool (Vlachos *et al*., 2014) in the DIANA-TarBase database (DIANA miRPath version 7.0) accessible at http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index.

This database (DIANA-miRPath) was used to determine the involvement and implication of the 5 miRNAs in cancer causing pathways and to further implicate them in the progression of PCa. The 5 miRNA were used as input into DIANA-miRPath. The microT-CDS tool was used to search for gene pathways, with a p-value threshold of 0.05 and micro-T threshold of 0.8 used as parameters. The resulting gene pathways were unionised and visualized as a heat map.

2.12.2 Prediction of Target Genes for Identified Prostate Cancer miRNAs

2.12.2 (a) miRNA Target Prediction with TargetScanHuman

Each of the five miRNAs designated miR1, miR2, miR3, miR4, miR5, were used as input in the TargetScanHuman search box using the miRNA nomenclature; hsa-miR-1 to hsa-miR5 for the respective miRNAs. Human was selected as species of priority from which to search for target genes. The queries were entered one at a time for each miRNA. TargetScanHuman returned a list of the

top 100 predicted target genes for each miRNA, irrespective of site conservation, ranked by cumulative weighted context++ scores for each miRNA. From each gene list obtained, only genes with a target prediction score of 80 % and above was selected for further analysis.

2.12.2 (b) miRNA Target Prediction with miRDB

Another online database used for target prediction was miRDB. As mentioned in section 2.3, two databases were used to improve the robustness of the predictions. Each of the five miRNAs was entered into the miRDB search box individually. From the drop down menu, human was chosen as the priority species from which targets were to be searched. The query was submitted using the 'Go' option. miRDB returned a list of predicted targets for each miRNA, from each gene list obtained, only genes with a target prediction score of 80 % and above were selected for further analysis.

Duplicates were removed from the individual gene lists obtained from the two platforms. These lists were combined and duplications were removed in Microsoft Excel. The result was a gene list of 502 genes. All the genes were represented by their official gene symbols

2.12.3 Functional Characterization of Predicted Genes via DAVID

The functional annotation of the 502 target genes identified was done using the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (https://david.ncifcrf.gov/summary.jsp) DAVID version 6.7 (Jiao *et al*., 2012). The gene list was copied and pasted into DAVID search box and the official gene symbol was selected as the unique gene identifier.. Annotations were limited by selecting for *Homo sapiens,* with the classification stringency set to medium. The "options" were set as follows; display, fold change and Bonferroni analysis. The list was then submitted for functional annotation clustering. Using the medium classification stringency for clustering, a total of 63 clusters were generated. Clusters generated were individually investigated to select clusters of genes that are involved in biological processes, present in the membrane region and that were involved in pathways that are known to play a critical role in the onset and progression of cancer. A total of 12 genes were produced from functional annotation in DAVID that corresponded to the selection criteria used.

2.12.4 Literature Review of Genes

Literature mining for each gene obtained from DAVID was performed. The following platforms (Uniprot, PolySearch, Google Scholar and GoPubmed) were used to search for abstracts or journal articles implicating the genes in cancer. Subsequent to these mining approaches, a final list of 9 putative genes was compiled. These were then cross-referenced back to the 5 miRNAs for their

involvement in prostate cancer. However, only 4 out of the 5 miRNAs had gene targets. The criteria used in selecting the final gene list after the enrichment in DAVID resulted in miR1 not being linked to a target gene. The gene list produced in DAVID was analysed for protein/gene expression analysis in STRING.

2.12.5 Analysis of Gene/Protein Interaction Networks via STRING

Gene IDs for the 9 genes, targeted by the 4 miRNAs implicated in prostate cancer were used as input for the generation of a gene network using the STRING DB Version 9. (Meiring, 2003; Franceschini *et al*., 2013). The 9 genes were used as driver genes to produce expression networks. To produce each of the expression networks, parameters were chosen as a follows: (i) a confidence level of 0.7, (ii) a network depth of 4 and (iii) restricting to show only the top 50 interactions between the 9 genes targeted by four miRNAs.

2.12.6 Analysis of Tissue-Specific Gene Expression Profiles via TiGER

Each of the 9 genes using their gene symbols was used as input into the "search box" option in the Geneview tool of TiGER to search for gene expression in prostate tissue.

2.12.7 Digital Expression Analysis via GeneHubGEPIS

The 9 genes were submitted by inputting the official gene symbol of each gene into the “search box” option, selecting ‘human’ as target species after which, selecting the “enter” option.

2.13 Results and discussion

2.13.1 Analysis of DIANA-TarBase generated pathways

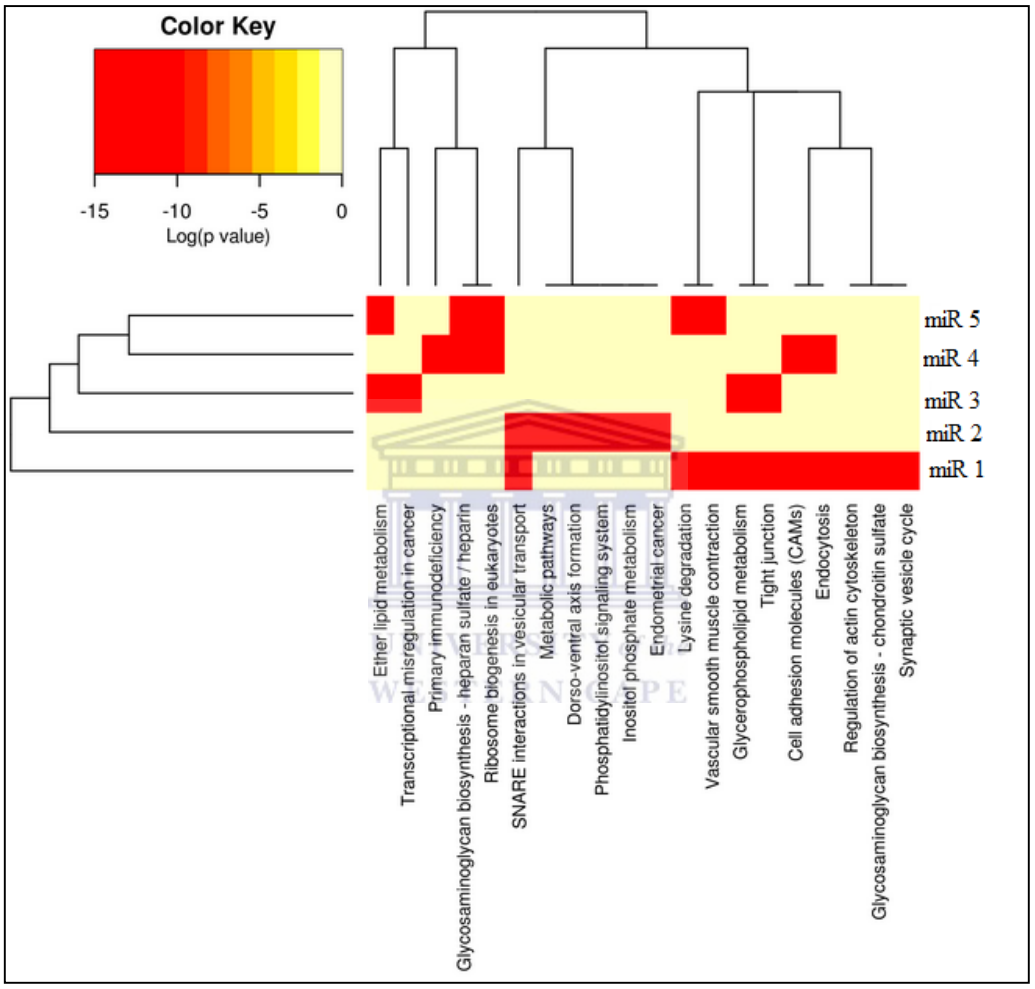


Figure 2.2 Association of the miRNAs in cancer causing related pathways. The strongest association is indicated by the red blocks. These have a log p-value -10≤. Intermediate association is indicated by the orange colour with a log p-value between -9.4 and -4.9. The weakest association is shown by yellow blocks with a log p-value between -1.5 and -5.

DIANA-TarBase aims to catalogue the published experimentally validated miRNA:gene interactions (Vlachos *et al*., 2014). The process of cataloguing miRNA targets faces major challenges as there are large numbers of interactions which the identified genes as well as the miRNAs are involved in (Xiao *et al*., 2009; Hsu *et al*., 2014). Thus an *in silico* approach is usually relied on to examine miRNA regulatory networks and their effects on molecular pathways.

Figure 2.2 shows the association of miRNAs in cancer causing related pathways. The analysis was done using the online publicly available tool DIANA-TarBase version7.0 (http://www.microrna.gr/miRPathv2). The figure represents a heat map of the 5 miRNAs. The analysis in DIANA showed that possible gene targets of miR1 are involved in cellular processes such as regulation of the actin cytoskeleton, glycerolphospholipid metabolism and lysine degradation.

It has been well documented that cancer cells reprogram their metabolic pathways to meet their abnormal demands for proliferation and survival (Tennant *et al*., 2010; Cairns *et al*., 2011). This is because they need a higher rate of metabolism to support their accelerated proliferation rate (Tennant *et al*., 2010; Cairns et al., 2011). Studies by Menendez and Lupu 2007; Cairns *et al*., 2011; Zhang and Du, 2012 have shown that altered lipid metabolism has been recognized as a common property of malignant cells. Because lipid metabolism in cancer cells is regulated by common oncogenic signalling pathways (Chunfa and Freter, 2012), and is believed to be important for the initiation and progression of tumours (Menendez and Lupu 2007), miR1, miR3 and miR5, predicted to be involved in this pathway, may be good indicators of the onset of cancer.

One of the predicted pathways that miR2 is involved in is the inositol phosphate metabolism pathway (Figure 2.2). Members of this pathway regulate cell proliferation, migration and phosphatidylinositol-3-kinase (PI3K)/Akt signalling (Yu et al., 2009). A recent study, (Tan *et al*., 2015) has shown that genes in this pathway are frequently dysregulated in cancer. The study found that gene dysregulation in this pathway was significantly associated with a risk of lung, breast, prostate and bladder cancers.

2.13.2 miRNA target prediction

Two publicly available target search databases, TargetScan and miRDB were used as platforms to identify genes associated with the five miRNAs.

Table 2.1: Representation of the number of target genes identified.

Database Name	# of target genes identified
miRDB	1076
TargetScanHuman	1502

The target genes obtained from the two databases were prioritized using the following criteria; (a) only the genes with a prediction score of 80 % and above were taken for further analysis, (b) gene lists with more than 800 targets were excluded, (c) duplications were eliminated in Excel.

Table 2.1 shows the number of target genes identified by the two different databases after applying the above criteria. The list from TargetScanHuman was 1502 genes and miRDB produced 1076 genes. (A combined list of gene targets from TargetScan and miRDB can be found in appendix A). There was a difference of 426 genes between the two databases used. This 28 % difference in the number of genes targeted by the miRNAs could be explained by the properties (target prediction algorithms) of both databases. As mentioned in section 2.5, miRDB avoids database redundancy (Wang, 2008; Peterson *et al*., 2014; Wong and Wang, 2015) and only considers seed matches in the 7-mer and 8-mer regions, whilst TargetScanHuman includes 6-mer regions in addition to 7-mer and 8-mer regions (Wang, 2008). The two gene lists were combined and duplicates were removed, resulting in a final gene list of 502 genes.

2.13.3 Functional Annotation via DAVID

The 502 genes prioritized in section 2.12.1 were used as input into DAVID for functional annotation and analysis. A classification stringency of medium in was employed in DAVID. The output was a total of 150 genes in 86 clusters. DAVID categorized the gene products into three Gene Ontology (GO) groups, namely; Cellular Component (CC) Biological Process (BP) and Molecular Process (MP), as seen in figures 2.3, 2.4 and 2.5 respectively.

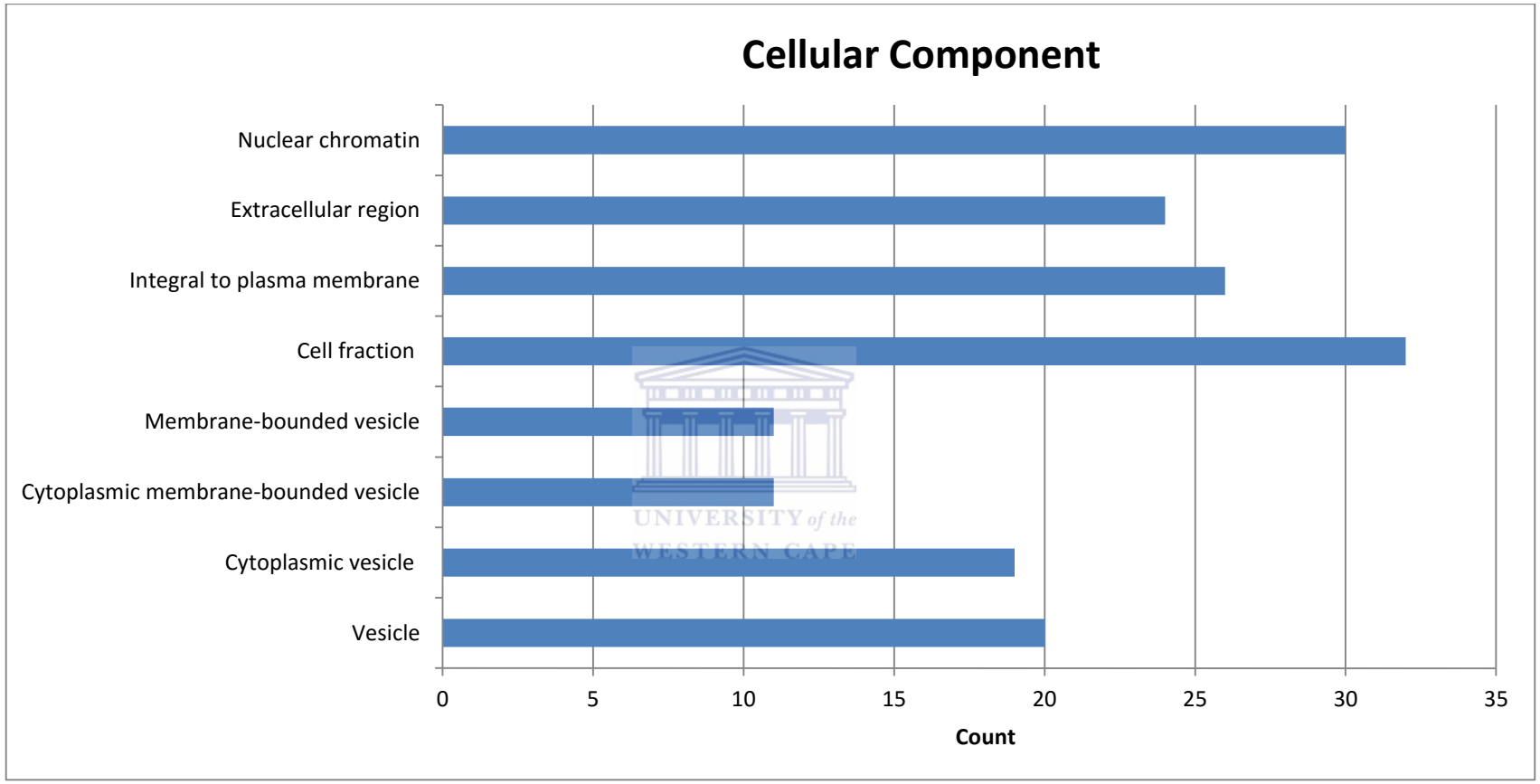


Figure 2.3: Functional characterizations of miRNA target genes under cellular component using DAVID. The blue bars represent the number of genes associated with the specified GO term.

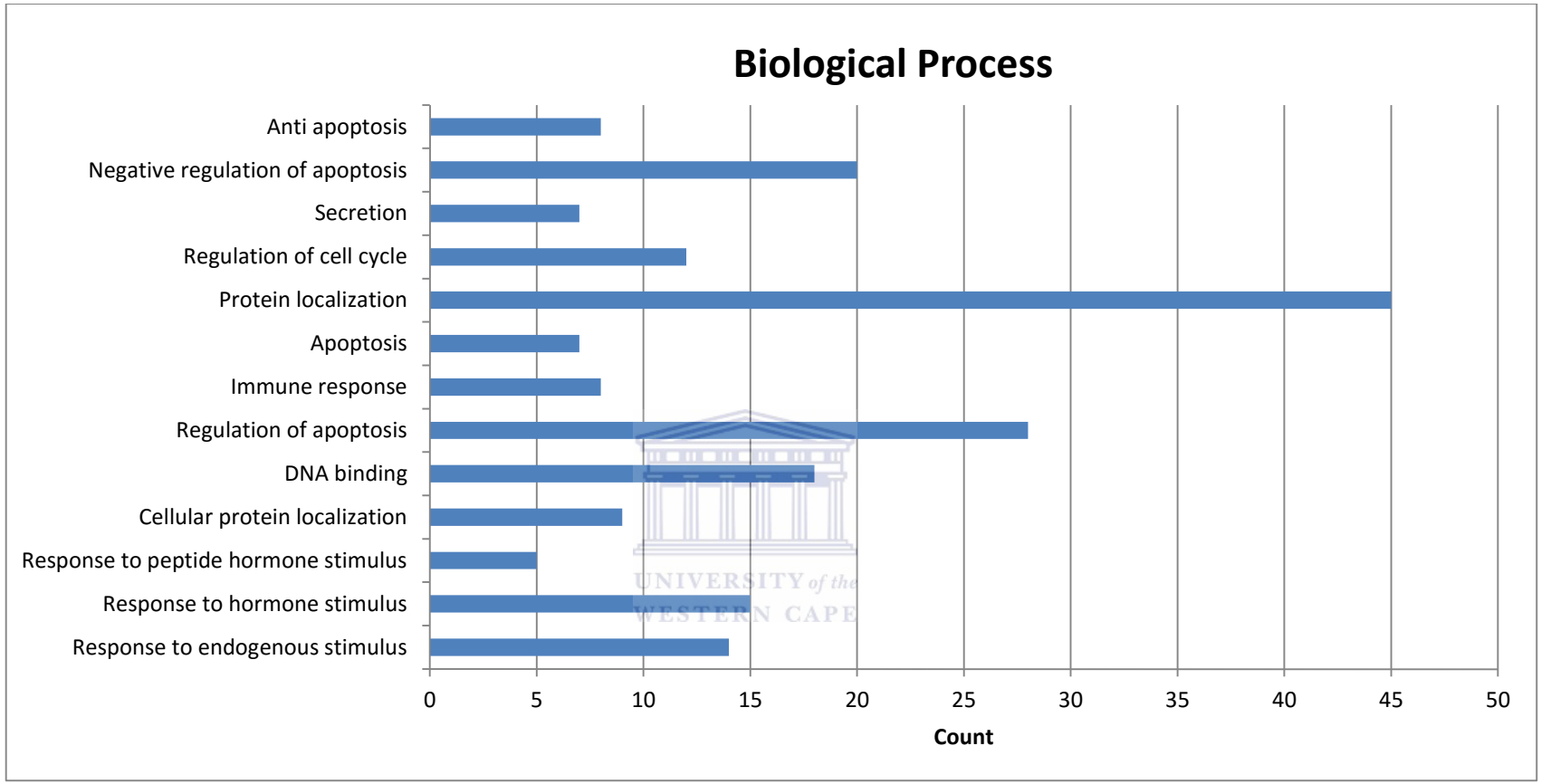


Figure 2.4: Functional characterizations of miRNA target genes under biological processes using DAVID. The blue bars represent the number of genes associated with the specified term.

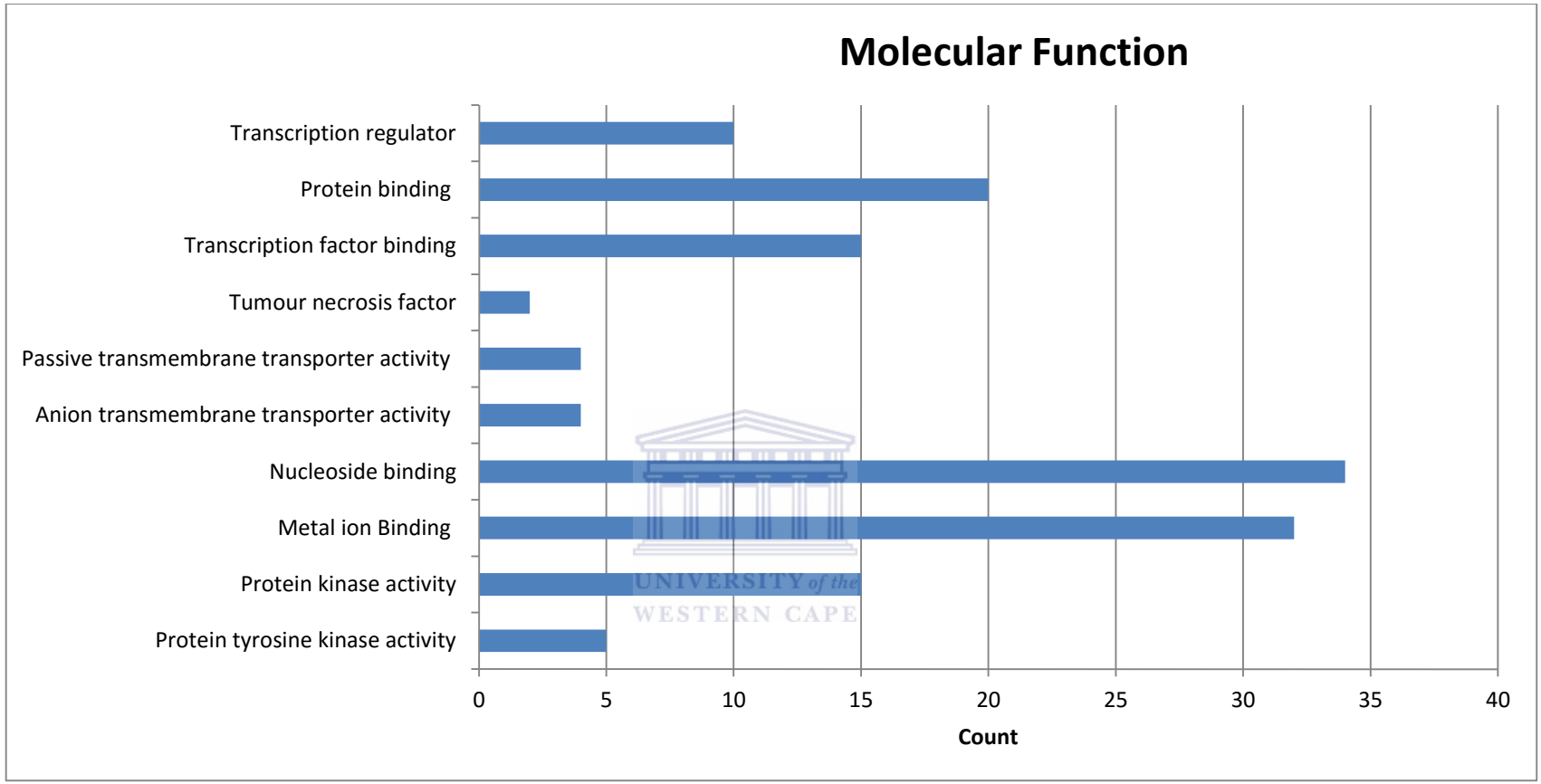


Figure 2.5: characterizations of miRNA target genes under molecular function using DAVID. The blue bars represent the number of genes associated with the specified GO term.

Table 2.2: miRNA target genes involved in the cellular component domain.

Gene	Cellular Component
MNT	-
AFAP1	Integral to plasma membrane, cell fraction
BIRC2	Cytoplasm membrane bound vesicle, cell fraction
LIG4	Integral to plasma membrane
YWHAZ	Cell fraction, vesicle, membrane bound vesicle,
TNFSF15	Extracellular region, integral to plasma membrane
TNFSF13B	Extracellular region, integral to plasma membrane, extracellular region
CTNND1	Cell fraction, membrane bound vesicle
FOXC1	Nuclear chromatin

Table 2.2 shows the breakdown of the 9 target genes involved in the cellular component. From figure 2.3, it can be seen that about 23 genes function in the extracellular region. Two of the target genes have been found to function in extracellular region, namely TNFSF13B and TNFSF15; they constitute 8.7 % of the total number of genes. The gene MNT did not have a cellular component function. AFAP1, TNFSF15, TNFSF13B and LIG4 constitute 15.4 % of the total genes characterized in DAVID. BIRC2 makes up 9 % of the total genes coding for proteins functioning as cytoplasm membrane vesicles (figure 2.3). YWHAZ and CTNND1 make up 6 % of the genes coding for proteins found in the cellular fractions. FOXC1 makes up 3 % of the total number of genes involved in nuclear chromatin functions.

Table 2.3: miRNA target genes involved in the biological process domain

Gene	Biological Process
MNT	Regulation of transcription, regulation of apoptosis, immune response
AFAP1	-
BIRC2	Response to endogenous cell death, apoptosis, negative regulation of apoptosis
LIG4	Negative regulation of apoptosis,
YWHAZ	Immune response, protein localization
TNFAF15	Negative regulation of apoptosis
TNFAF13	Negative regulation of apoptosis, immune response,
CTNND1	Protein localization
FOXC1	Regulation of apoptosis

Table 2.3 shows the breakdown of the 9 target genes involved in the biological process domain. The biological process ontology constituted the greater number of the gene functions. The genes MNT and FOXC1 made up 7 % of the total genes involved in regulation of apoptosis. BIRC2, LIG4, TNFAF15 and TNFAF13B made up 20 % of genes involved in negative regulation of apoptosis. The genes YWHAZ and CTNND1 constituted 4 % of the genes involved in protein localization (figure 2.4). The gene AFAP1 did not have any predicted biological process function.

Table 2.4: miRNA target genes involved in the molecular function domain.

Gene	Molecular Function
MNT	Transcription factor binding, transcription regulator
AFAP1	Protein binding
BIRC2	Metal ion binding
LIG4	Metal ion binding, nucleoside binding
YWHAZ	Transcription factor binding
TNFSF15	Tumour necrosis factor
TNFSF13B	Tumour necrosis factor
CTNND1	Protein binding
FOXC1	Transcription regulator

Table 2.4 shows the breakdown of the 9 target genes involved in the molecular function domain of the gene ontology. It can be seen from figure 2.5 that 20 genes are involved in protein binding. Two of the target genes are found to be represented in that number, namely CNTDD1 and AFAP1. FOXC1 and MNT constitute two out of 10 of the transcription regulator genes. The genes TNFSF15 and TNFSF13B are the only genes that code for the tumour necrosis factor. LIG4 and BIRC2 represent 66 % (2 out of 30) genes involved in metal ion binding (figure 2.5). The gene YWHAZ constitutes 6.7 % of the genes coding for transcription factor binding.

Table 2.5: Representation of the miRNAs and their identified gene targets from DAVID.

miRNA	Confidence %	Target (Gene)	Gene description
miR4	96	TNFSF15	Tumour necrosis factor (ligand) superfamily, member 15
miR4 **miR3**	80	LIG4	Ligase IV, DNA, ATP-dependent
miR4	99	FOXC1	Forkhead box C1
miR5	97	YWHAZ	Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein, zeta polypeptide
miR5	88	TNFSF13B	Tumour necrosis factor
miR4 **miR2**	98 83	AFAP1 (AFAP)	Actin Filament Associated Protein
miR4	83	CTNND1	(E-cadherins) cell malignancy
miR4	87	BIRC2	Apoptotic suppressor
miR3	80	MNT	MAX binding protein

One feature of the functional annotation tool in DAVID is to give predictions on the localisation of proteins encoded by the input genes. For purposes of this study, emphasis was placed on genes with proteins localised on the cell surface. Figure 2.3 depicts miRNA targets strongly associated with membrane and membrane bound proteins. In the process of tumour development, or invasion, some proteins involved in cell-cell adhesion or cell movement lose their adhesive properties and are shed into the surrounding environment (Hanash, 2011) finding their way into biological fluids, such as saliva, blood and urine. This makes them

good targets as biological markers for therapeutics and diagnostics. Other important aspects that were taken into consideration when prioritizing the gene list were the processes they are involved in. Genes involved in biological processes such as regulation of apoptosis, secretion and immune response were prioritized, figure 2.4. These genes were considered significant as apoptosis is important in cancer development. Studies have shown that some oncogenic mutations disrupt apoptosis, leading to tumour initiation, progression or metastasis (Lowe and Lin, 2000). Conversely, compelling evidence indicates that other oncogenic changes promote apoptosis, thereby producing selective pressure to override apoptosis during multistage carcinogenesis (Su *et al*., 2015).

Figure 2.5 shows that 30 % of the target genes are involved in transporter membrane activity. Recent studies have shown that plasma membrane transporter proteins play an important role in taking up nutrients into and effluxing xenobiotics out of cells to sustain cell survival (Williams, 2014). In the last decade, it has been reported that many transporters are differentially up-regulated in cancer cells compared to normal tissues, suggesting that the differential expression of transporters in cancer cells may become a good target for diagnostic markers for cancer therapy.

The outcome of functional annotation using DAVID, coupled with literature mining was a list of 9 genes. A number of these genes, depicted in table 2.5, have been implicated in cancer (YWHAZ, FOXC1 and TNFSF13B) as well as prostate cancer, (MNT) (van Rooij, 2011; Dedeoğlu, 2014; Stokowy *et al.,* 2014; Uso *et al.,* 2014; Dedeoğlu, 2014). The remaining genes do not appear in literature as having been validated as prostate cancer biomarkers or connected to the disease.

2.13.4 Gene/Protein Interaction Analysis via STRING

The 9 genes were used as input into STRING Version 9 to visualize protein expression networks as described in section 2.12.5 Figures 2.4 and 2.5 show the results of the analysis. Figure 2.12.5 shows that there was an association between TNFSF15 and TNFSF13B which code for the human tumour necrosis factor ligand superfamily member proteins, which are also known as proliferation-inducing ligands. Both these two genes may play a key role in the development of B-cells (Boss *et al*., 2006) and plasma cells in the bone marrow (Matthes *et al*., 2011). They also function in the development of human tumours (Zhao *et al*., 2014). TNFSF13B expression is related to the progression of several types of carcinomas (Moreaux *et al*., 2009; Pelekanou *et al*., 2011) including renal cell cancer, (Pelekanou *et al*., 2011), breast cancer, (Zhao *et al*., 2014; Moreaux *et al*., 2009) and lung cancer (Lin *et al*., 2012).

It can be seen in figure 2.4 that TNFSF15 and TNFSF13B are both linked to LIG4 by the apoptotic suppressor BIRC2. BIRC2 is a member of the inhibitor of apoptosis family of proteins and plays a pivotal role in regulation of nuclear factor-κB (NF-κB) signalling and apoptosis (Gyrd Hansen *et al*., 2010). Up-regulation of BIRC2 has been frequently detected in lymphoid malignancies (Gyrd Hansen *et al*., 2010). A recent study (Yamato *et al*., 2015) showed that up-regulation of BIRC2 is evident in a wide range of epithelial tumours such as lung tumours.

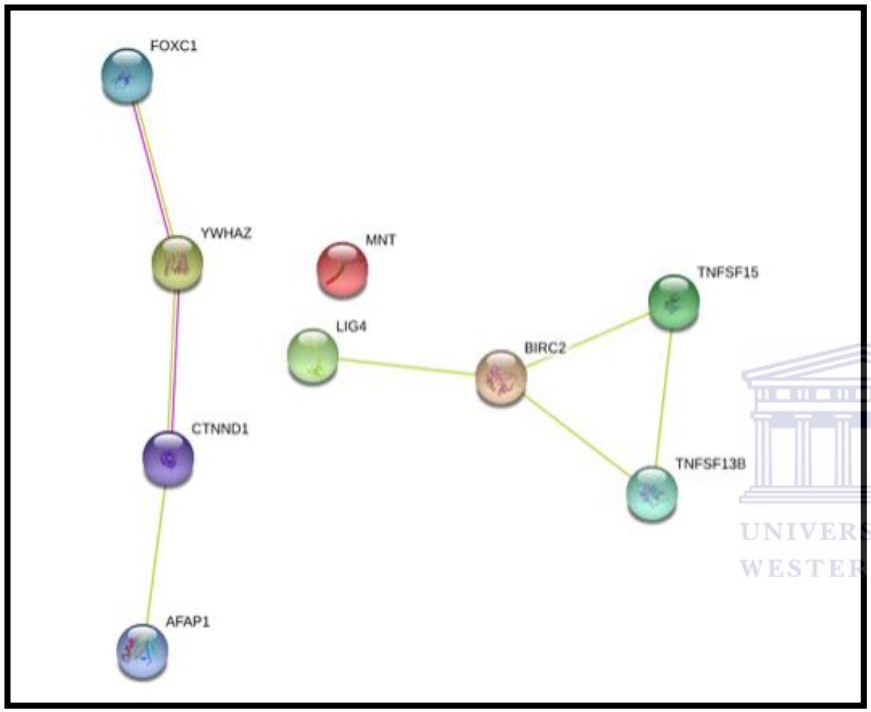


Figure 2.6: Protein Network Visualization generated by STRING. The interactions of the 9 miRNA targeted genes clustered together. The genes are represented by the nodes and the different line colours represent the types of evidence for the association.

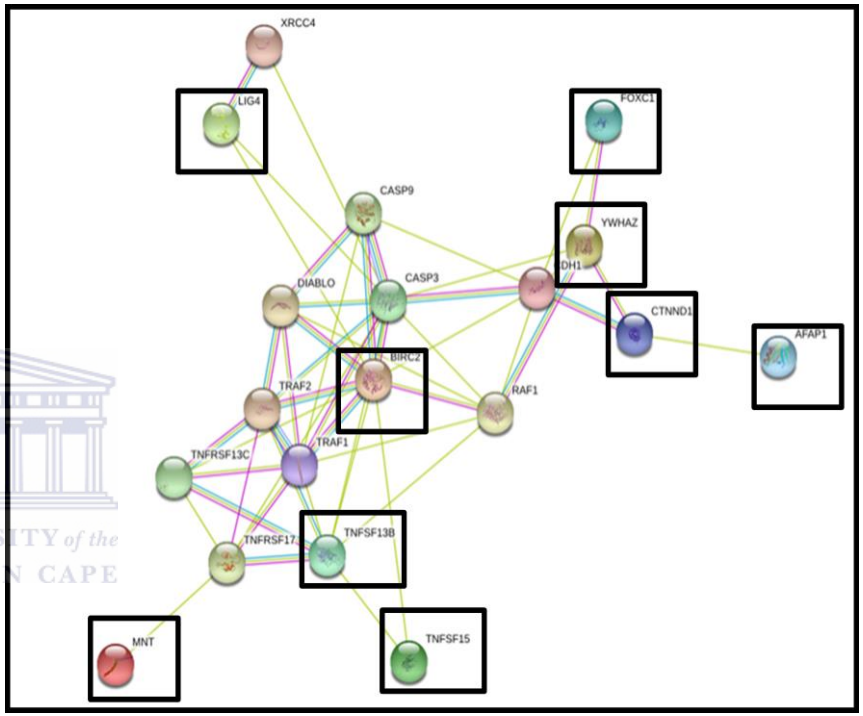


Figure 2.7: Protein Network Visualization generated by STRING. The interactions of the 9 miRNA targeted genes in in association with other genes. The genes are outlined in the black boxes.

The gene LIG4 encodes the protein DNA ligase that joins single-strand breaks in a double-stranded polydeoxynucleotide in an ATP-dependent reaction (Francis *et al*., 2014; Xie *et al*., 2014). Thus this protein is essential for recombination and DNA double-strand break repair. Dys-regulation of this gene by miRNAs may influence DNA repair ability, thus altering genetic stability and resulting in carcinogenesis. The genes AFAP1, CTNND1, YWHAZ and FOXC1 are linearly connected. AFAP1 is thought to encode a protein that is a potential modulator of actin filament integrity which responds to cellular signals. One of the Hallmarks of cancer is cell signalling, with cancer cells being self sufficient with regards to needing external signals to grow and divide, or insensitivity to anti-growth signals (Weinberg and Hanahan, 2000; Hanahan, 2011). Perhaps dys-regulation in this gene may result in cell signalling insufficiency or insensitivity. However, further studies would have to be done to prove this.

The FOXC1 gene belongs to the forkhead family of transcription factors which is characterized by a distinct DNA-binding forkhead domain. The specific function of this gene has not yet been determined. However, a recent study implicated it in brain metastasis in breast cancer (Ray *et al*., 2010; Sizemore *et al*., 2012). The MAX binding protein MNT did not have any connections to the other genes.

Figure 2.7 shows the predicted interactions of the 9 miRNA target genes with each other as well as with other genes. The 9 genes of interest are outlined in solid boxes. Thus, the 9 candidate genes are involved in metabolic and cellular pathways with not only each other but with other genes. The associations are based on the co-occurrence of proteins on metabolic maps in the KEGG database (von Mering *et al*., 2003). Proteins that occur on the same metabolic KEGG map

are presumed to be functionally interacting. The different coloured lines represent evidence for the association. STRING, displays a score of confidence for each association between proteins (low confidence: scores <0.4; medium: 0.4 to 0.7; high: >0.7). The higher the score, the more evidence for the association recorded by STRING (Eisen *et al*., 1998; von Mering *et al*., 2003; Franceschini *et al*., 2013). Some genes have more than one association. They have a combined score which is close to one. Thus, the more combined associations between genes, the more evidence for the interaction. (Franceschini *et al*., 2013). Upon visualization of predicted interaction networks of BIRC2 (figure 2.7), it can be seen that it has many associations. This shows that it is an important protein to the network. The more interactions a protein has in a network, the more important that protein is to the network as many other proteins rely on it in their functioning (Tomaic *et al*., 2008). BIRC2 is connected to CASP3, DIABLO, RAF1, TRAF1 and TRAF2. It is connected to these genes by 3 lines of evidence, indicating more associations and thus higher confidence (Tomaic *et al*., 2008). The family of TRAF genes encode proteins that are receptor associated factors. They are involved in the regulation and response to apoptosis (Inoue *et al*., 2000; Potter *et al*., 2007). DIABLO is a mitochondrial protein that potentiates some forms of apoptosis (Adrain *et al*., 2001; Martinez-Ruiz *et al*., 2008). CASP3 encodes proteins in the family of caspases. These are endoproteases that provide critical links in cell regulatory networks controlling inflammation and cell death. Thus CASP 3 is an apoptosis related cysteine peptidase (Mcllwain *et al*., 2013). Studies by Soung *et al*., 2004 and Chen *et al*., 2008 found that alterations in this gene may lead to human tumourigenesis.

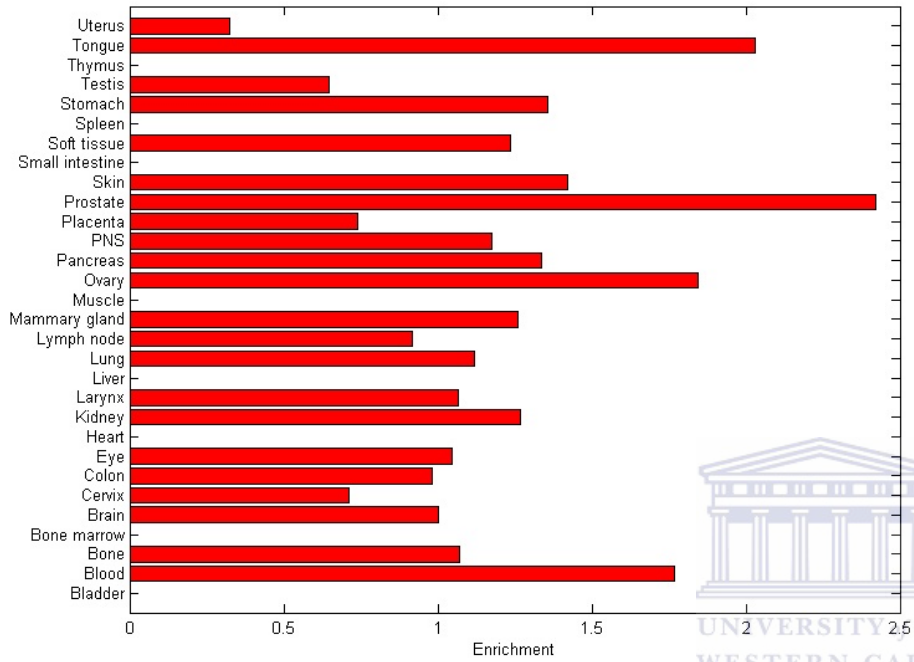
Upon visualization of predicted interaction networks of YWHAZ, it was observed that it has two interactions with FOXC1, two with CTNND1, two with RAF1 and one interaction with CASP3. This is a total of 7 interactions. This shows that it is an important protein, central to this particular network.

2.13.5 Tissue Specificity Expression Analysis via TiGER and GeneHub-GEPIS

The 9 genes were subjected to cross cancer tissue specific analyses in TiGER and GeneHub-GEPIS (section 2.12.6-7). Figures 2.8 to 2.10 give a graphical display of 3 genes and their expression in prostate tissue for both databases. Data on the remaining 6 genes can be found in Appendix A. GeneHUB-GEPIS database compares expression of the gene between cancerous and normal tissue.

A detailed understanding of how genes are expressed and regulated in different tissues can help elucidate the molecular mechanisms of tissue development and function. The approximately 25,000 genes (Liu *et al*., 2008) in the human genome demonstrate dramatic diversity in terms of expression levels, both temporally and spatially (Bartel, 2009). Despite this diversity, the expression of all genes is controlled by a relatively small number (less than 2,000) of transcription factors (Liu *et al*., 2008). Thus gene expression regulation also depends heavily on miRNAs (Bartel, 2009). Therefore, a database dedicated to comprehensive information about tissue-specific gene regulation is a desirable tool in expression studies.

The TiGER results for the MNT gene show that it is preferentially expressed in prostate tissue (Figure 2.8 A). The GeneHub-GEPIS results for MNT show a DEU value of 94.6 in over-expressed cancerous prostate cells when compared to the normal prostate tissue (Figure 2.8 B). This figure is higher than the normal tissue expression and MNT's over-expression in any other tumour tissue. This gene is targeted by miR3, thus it is a potential biomarker for prostate cancer detection. MNT codes for a protein that is a transcriptional repressor and an antagonist of Myc-dependent transcriptional activation and cell growth (Montagne *et al*., 2008). This could explain its over-expression in prostate cancer tissue.



B

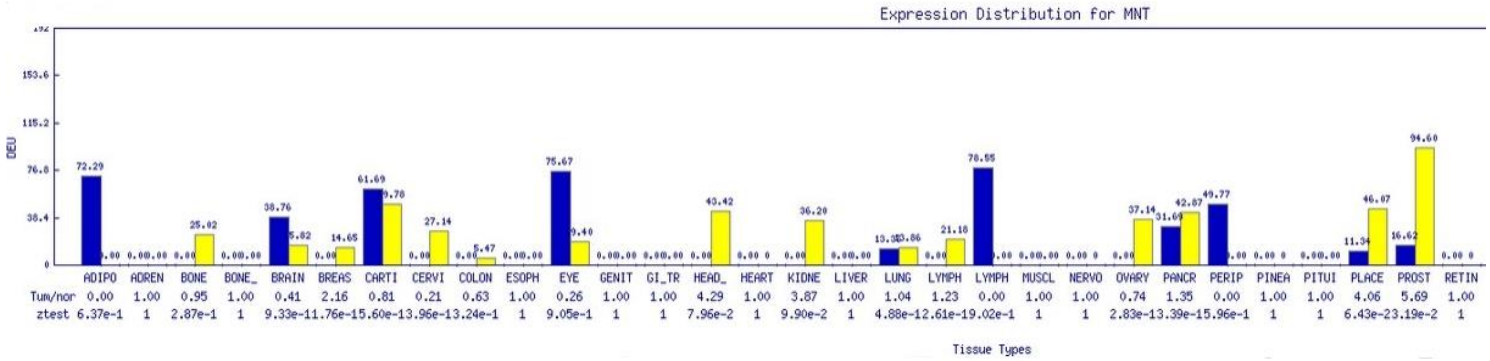


Figure 2.8: Expression profile for MNT from TiGER (A) MNT is preferentially expressed in the prostate with an expression enrichment value greater than 2. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

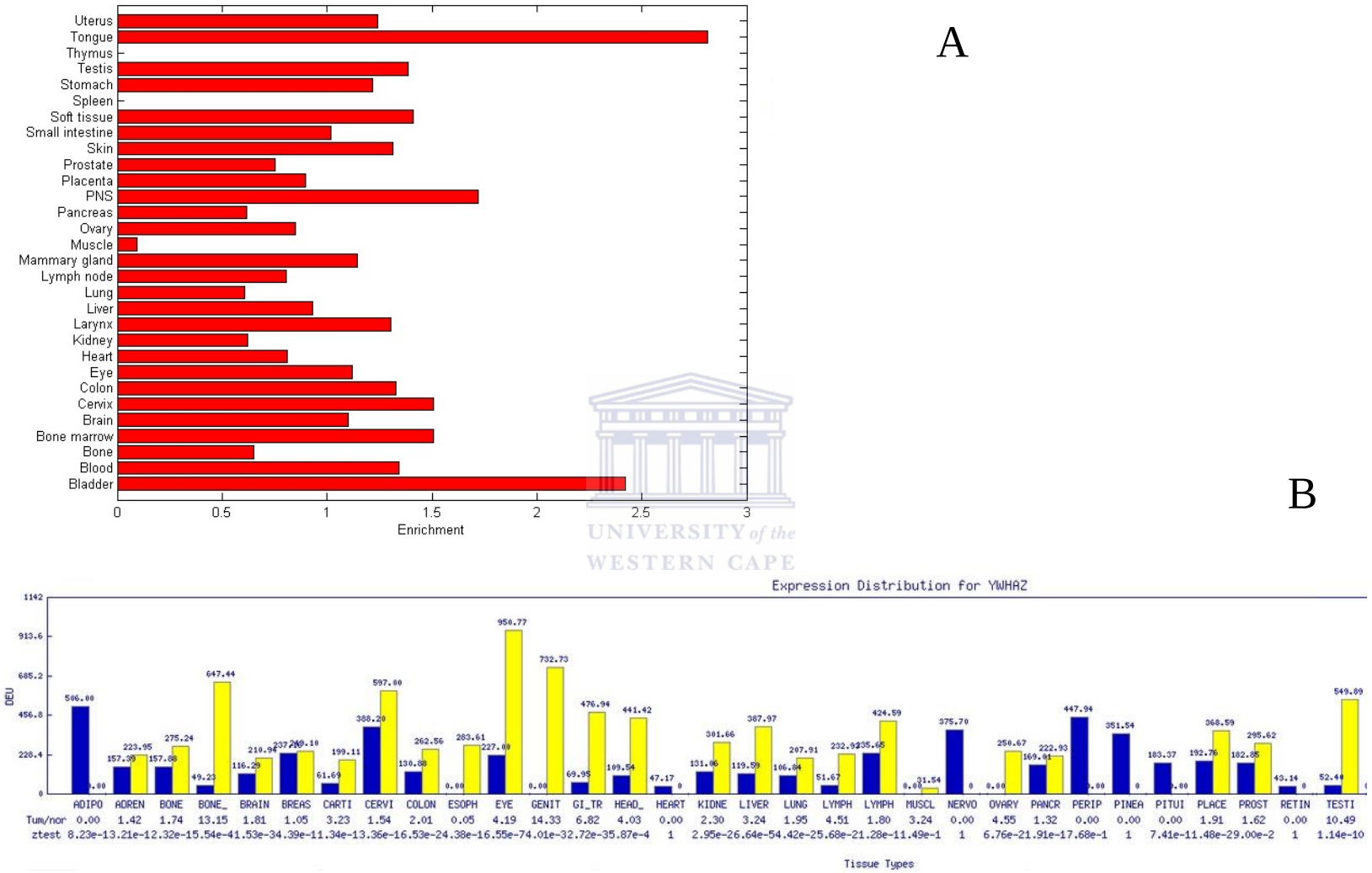
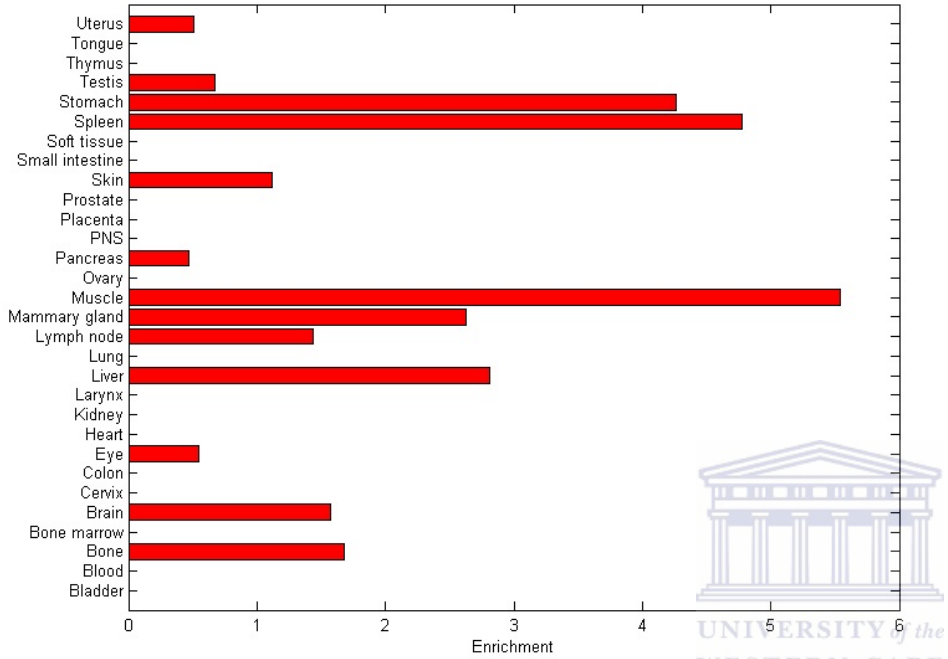


Figure 2.9: Expression profile for YWHAZ from TiGER (A) YWHAZ is preferentially expressed in the tongue with an expression enrichment value greater than 2.5. Expression in prostate tissue is between 1 and 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

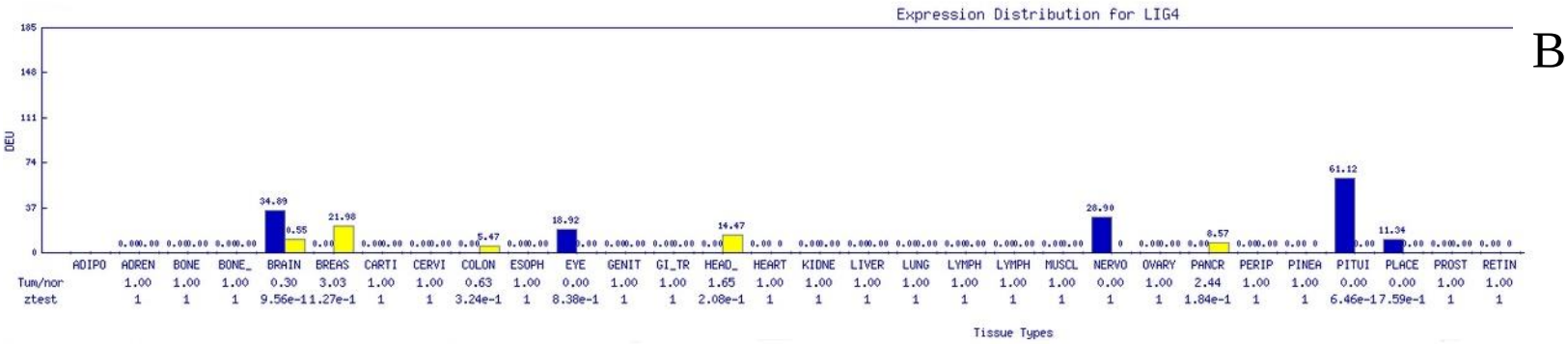
Figure 2.9 show expression profiles of the gene YWHAZ, targeted by miR5. TiGER results indicate that it is preferentially expressed in the tongue followed by the bladder with an expression enrichment of about 2.8 and 2.4 respectively. The GeneHub-Gepis results show a DEU of 588.53 in prostate tumour tissue. Thus, YWHAZ is over-expressed in prostate tumour when compared to normal prostate tissue. However, over-expression in eye tumours is greater at 950.77 DEUs. Therefore YWHAZ may not be a good biomarker specific to prostate cancer detection, further experimental evaluation is required.

There was no data present for AFAP and TNFSF15 in TiGER, figures 2.15 and 2.16 (appendix A). However, there was data on tumour and normal tissue expression for AFAP1 in GeneHub-GEPIS. The expression profile in GeneHub-GEPIS showed that it is under-expressed in prostate tumours when compared to the normal prostate tissue.

The lack of data for AFAP1 and TNFSF15 in TiGER could be as a result of lack of data for the particular genes in the database (not including the current data) or it could indicate that gene expression in tissues is not well documented for both the genes. The TNFSF15 gene targeted by miR2 and miR4 belongs to the tumour necrosis factor ligand family, as discussed earlier (section 2.13.3). It acts as an autocrine factor to induce apoptosis in endothelial cells (Michelsen *et al*., 2009). It has been reported to be abundantly expressed in endothelial cells. Thus, dysregulation can result in any number of cancers.



A



B

Figure 2.10: Expression profile for LIG4 from TiGER (A) LIG4 is preferentially expressed in muscle tissue with an expression enrichment value of about 4.5. There is no expression recorded in prostate tissue. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

From the results of the expression profiling in TiGER and GeneHub-GEPIS, there is evidence that genes targeted by the 5 miRNAs are expressed in prostate tissues. However, some of the genes are preferentially expressed in other tissues, with some genes such as LIG4 (figure 2.10) not having any expression at all in neither the normal prostate tissue nor the prostate tumour tissue. There could be three reasons for this, the gene is not expressed in the prostate tissue, the databases used are lacking information on this gene or thirdly, there is currently no experimental validation of LIG4 expression in prostate tissue.

2.14. Conclusions and Summary

It has been shown that miRNAs are involved in carcinogenesis (Croce *et al*., 2008) via regulation of gene expression (de Bock *et al*., 2010). Recognition of miRNAs that are differentially expressed between tumour tissues and normal tissues may help to establish the apparent pathogenic role of miRNAs in cancers (Lu *et al*., 2005; Karube *et al*., 2010; Zhang *et al*., 2007). Investigations (Croce *et al*., 2008) have demonstrated that almost all cancers have alternative miRNA expression profiles when compared to their adjunct normal tissues. These cancer types include lung cancer, leukaemia, brain cancer, prostate and breast cancer (de Bock *et al*., 2010), which together have caused the majority of cancer related deaths in the past decades (CDC, 2014).

In this study, putative miRNA biomarkers for prostate cancer were identified. A combination of biological data mining, text mining and *in silico* gene enrichment

techniques proved to be effective in classifying genes targeted by miRNAs and linking them to cancer. An analysis of predicted miRNA pathways in DIANA showed that they are involved in metabolic processes that that are important in cancer progression. Thus, it can be investigated further if they are involved in prostate cancer via qRT-PCR expression analysis.

The gene targets of the miRNAs annotated in DAVID showed that they are involved in the cellular component, biological processes and molecular function gene ontology domains. In these domains it was found that the genes code for proteins involved in various processes and cell functions whose dys-regulation could lead to cancer. Some of the genes such as MNT and FOXC1 were found to code for proteins that are involved in regulation of apoptosis. LIG4, AFAP1 and CTNND1 code for proteins integral to the plasma membrane. BIRC2, TNFSF13B and TNFSF15, were found to code for proteins that are involved in negative regulation of apoptosis. YWHAZ codes for genes involved in immune response.

STRING results indicated that the genes are involved in interaction networks with each other as well as other genes. It was noted that the genes are connected to CASP3, DIABLO, RAF1, TRAF1 and TRAF2 which have demonstrated roles in apoptosis. It was also found that BIRC2 is the most connected gene in the gene list. Thus it is an important part of the gene network as it connects many genes via its associations.

Tissue specificity expression analysis in TiGER revealed that MNT is preferentially expressed in prostate tissue. Results from GeneHUb-GEPIS showed that the gene was over expressed by a DEU of 94.6 in prostate tissue when compared to the normal tissue. The analysis also showed that, YWHAZ,

FOXC1 and TNFSF13B, were all found expressed in prostate tissue. LIG4 and TNFSF15 did not express in prostate tissue. AFAP1 did not show expression in prostate from the TiGER database. However, GeneHub-GEPIS showed that it is under-expressed in prostate tumours when compared to the normal prostate tissue.

Chapter 3

3. Molecular Validation of miRNAs as Putative Biomarkers for the Early Detection of Prostate Cancer

3.1 Introduction

It is well documented that miRNAs are involved in gene regulation (Lagos-Quintana *et al*., 2001; Bartel, 2004; Ardekani and Naeini, 2010). This is done through binding to target mRNA, thereby altering protein expression. As discussed in section 1.6.3, the binding of miRNA to the mRNA can inhibit protein translation or accelerate the process of deadenylation causing an mRNA to be degraded (Melo and Melo, 2014). Because of their function, miRNAs have played a vital role in disease pathogenesis and have potential as biomarkers and therapeutic agents (MacFarlane and Murphy, 2010; Yang *et al*., 2011; Worringer *et al*., 2014).

Currently, routine diagnostic methods for the early detection of PCa include digital rectal examination (DREs) and prostate-specific antigen (PSA) testing (American Joint Committee on Cancer, 2012; CANSA, 2015). The PSA test is nonspecific, due to the fact that elevated PSA levels have been measured in benign prostatic hyperplasia (BPH), infection, and/or chronic inflammation (Heidenreich *et al*., 2011; Basch *et al*., 2012). Thus, testing for PSA may lead to confounding outcomes. Other blood based biomarkers including the human

glandular kallikrein 2 (hK2) and urokinase plasminogen activator (uPA) and its receptor (uPAR), have been studied alone or in combination with PSA and suggested for diagnosis, staging, prognostics, (Karazanashvili and Abrahamsson, 2003; Ishiguro *et al.,* 2009; Nguyen *et al.,* 2014) and monitoring of prostate cancer (Nguyen *et al.,* 2014). However, because of the diverse nature of prostate cancer and its presentation in patients (Karazanashvili and Abrahamsson, 2003), there is an urgent need to identify additional biomarkers for enhanced prediction of disease progression and prognosis to aid in clinical decision making with respect to treatment options.

miRNAs have many required features of good biomarkers, they are stable in various bodily fluids, the expression of some miRNAs is specific to tissues or biological stages (Sita-Lumsden *et al*., 2013; Moldovan *et al*., 2014), and the level of miRNAs can be easily assessed by various methods of gene expression profiling such as the Polymerase Chain Reaction (PCR) (Derveaux *et al*., 2010). The changes of several miRNA levels in plasma, serum, urine, and saliva have already been associated with different diseases (Hanke *et al*., 2009; Fabbri 2010). For example, the ratio of miR-126 and miR-182 in urine samples can be used to detect bladder cancer (Hanke *et al*., 2009), and decreased levels of miR-125a and miR-200a in saliva is associated with oral squamous cell carcinoma (Ng *et al*., 2009).

Gene expression analysis is increasingly important in various biological research fields and understanding the patterns of expressed genes is expected to provide insight into complex regulatory gene networks and can lead to identification of

genes implicated in disease. Quantifying gene expression levels can yield valuable clues about the function of a gene, for instance, accurate measurements of gene expression can identify the type of cells or tissues where particular genes are expressed, reveal individual gene expression levels in defined biological states and detect alterations in gene expression levels in response to specific biological stimuli (Fraga *et al*., 2008).

3.1.1 Quantitative real-time PCR (qRT-PCR)

The quantitative real-time polymerase chain reaction (qRT-PCR) has recently become the most widely used technique in modern molecular biology (Derveaux *et al*., 2010). This technique depends on the fluorescence-based detection of amplicon DNA and permits the kinetics of PCR amplification to be monitored in real time, making it possible to quantify nucleic acids with ease and precision (Guescini *et al*., 2008). Quantitative real-time PCR (qRT-PCR) has become a very versatile technique to examine expression changes of one or more genes of interest in various pathological states such as cancer. This method offers a broad range of advantages over standard methods such as the Northern blot and semi-quantitative PCR due to its specificity, sensitivity, simplicity, low cost and high-throughput nature (Derveaux *et al*., 2010).

3.1.2 Quantification Strategies in qRT-PCR

There are two strategies used to quantify gene expression in qRT-PCR; absolute quantification and relative quantification. Absolute quantification relies on a standard curve which is generated by using a serially diluted sample of known concentration. The log of these dilution concentrations are plotted against the crossing points (Cp).

The crossing point (Cp) is the cycle at which fluorescence achieves a defined threshold (Taylor, 2010). It corresponds to the cycle at which a statistically significant increase in fluorescence is first detected (Pabinger *et al*., 2009). The Cp value decreases linearly with an increase in target quantity. Thus, Cp values can be used as a quantitative measure of the input target number (Heid *et al*., 1996).

Relative Quantification compares the levels of two different target sequences in a single sample, such as the target gene of interest and another gene; and expresses the final result as a ratio of these targets (van Peer *et al*., 2012). For comparison purposes the second gene is a reference gene that is found in constant copy numbers under all test conditions (Yuan *et al*., 2006). This reference gene, which is also known as endogenous control, provides a basis for normalizing sample-to-sample differences (Heid *et al*., 1996).

3.1.3 Aims and Objectives

This section of the study aimed to validate the 5 miRNAs identified by *in silico* expression analysis in chapter 2 using molecular methods. Specific objectives are outlined as follows;

1.) Growth of a prostate cancer cell line, a benign prostate hyperplasia cell line, other cancer cell lines as well as non-cancerous cell lines including a normal prostate cell line from which miRNA was extracted.
2.) Extraction of miRNA and synthesis of cDNA
3.) Molecular expression profiling of the 5 miRNA (cDNA) via Real-time Polymerase Chain Reactions (qRT-PCR) in all cell lines.

3.2 Materials and Methods

Reagents and suppliers.

Table 3: General reagents and suppliers

Reagent	Supplier
Dulbecco's Minimal Essential Medium (DMEM)	Lonza
Dulbecco's Minimal Essential Medium (DMEM) F12	Lonza
Roswell Park Memorial Institute Medium (RPMI) 1640	Lonza
Dimethyl Sulphoxide (DMSO)	Sigma
Fetal Bovine Serum (FBS)	Lonza
Phosphate Saline Buffer (PBS)	Lonza
Trypsin	Lonza
β-mercaptoethanol	Sigma
KAPA SYBR® FAST qRT-PCR Kit	KAPABiosystems
RNA isolation kit Nucleospin®	Macherey-Nagel
Nuclease free dH_2O	Merck

3.2.1 Cell Culture

Cell lines used in the study are depicted in table 3.1. The cell lines that were used in the course of this research were purchased from American Type Culture Collection (ATCC). The PNT1a cell line (passage 4) was obtained from Luiz Zerbini of the International Centre for Genetic Engineering and Biotechnology Cape Town (ICGEB). All cell lines are epithelial and adherent except the BPH1 which is epithelial but non-adherent.

Table 3.1: Cell lines used to investigate the specificity of the miRNAs in cancer.

Cell line	Tissue	Description	Growth medium
PNT1a	Prostate	Non-cancerous prostate epithelial cell line	RMPI, PenStrep 10 % FBS
LNCaP	Prostate	AR-sensitive prostate carcinoma	RPMI, PenStrep 50 % FBS
BPH1	Prostate	Benign prostate hyperplasia	RPMI, PenStrep 10 % FBS
A549	Lung-Alveolar	Lung epithelial carcinoma	DMEM, PenStrep 10 % FBS
HEPG2	Liver	Hepatocyte epithelial carcinoma	DMEM, PenStrep 10 % FBS
MCF12	Breast	Non-cancerous Breast cell line	DMEM/F12, PenStrep 10 % FBS, Hdyrocortisone, EGF, Insulin
MCF7	Breast-pleural effusion	Breast adenocarcinoma	DMEM, PenStrep 10 % FBS
H157	Buccal Mucosa	Squamous cell carcinoma	DMEM, PenStrep 10 % FBS
KMST6	Fibroblast	Non-cancerous embryonic fibroblast	DMEM, PenStrep 10 % FBS
HeLa	Cervix	Cervical adenocarcinoma	DMEM, PenStrep 10 % FBS
HT29	Colon	Colorectal adenocarcinoma	DMEM, PenStrep 10 % FBS

3.2.2 Start up of Cell Culture from frozen Cells

The frozen cryovials were held under 25 ºC running tap water for about one minute until defrosted. The vial was wiped down with 70 % ethanol and placed in a lamina flow hood where the vial contents were emptied into a 15 mL tube to which 5 mL of pre-warmed complete medium was added (table 3.1). The tube was then centrifuged for 5 minutes at 2039 x g using a Sorvall H4000 TC6 centrifuge (American Instrument Exchange, Inc). The supernatant was removed and discarded. The pellet was re-supended in fresh culture medium and the suspension transferred to a 25 cm^2 flask (T25). The flask was then incubated in humidified incubator at 37 ºC with 5 % CO_2 for 24 hours after which the medium was checked for contamination visually, this would be indicated by cloudy medium . After additional 24 hours of culturing, the flask was viewed under a Nikon TMS microscope at a magnification of 200 X to check if the cells had adhered to the flask. The medium was removed and replaced with fresh culture medium to remove any traces of DMSO left over from the cryopreservation medium.

3.2.3 Maintaining the cell lines

A schedule of cell culture maintenance was conducted as follows; medium was visually examined for contamination daily and flasks were examined under a microscope for culture confluence. When contamination was not observed and the confluency was below 50 %, old medium was replaced with fresh complete growth medium. At a confluency of 80 %, the cells were sub cultivating.

3.2.4 Subcultivation

3.2.4.1 Adherent cultures

To sub-cultivate (passage) the cultures, old medium was aspirated with a sterile Pasteur pipette and discarded. The culture was then washed with 3 mL 1X Phosphate Buffered Saline (PBS) (table 3) pre-warmed at 37 ºC. The flask was swirled for 15 seconds after which the PBS was aspirated with a Pasteur pipette and 1 mL of 1.25 % trypsin (table 3) was added to the culture. The flask was swirled to spread the trypsin evenly. The flask was then placed in an incubator at 37 ºC for 2 minutes after which the culture was viewed under a microscope to check for detachment of the cells. When the cells were detached, 5 mL of fresh complete growth medium was added to deactivate the trypsin. The suspended cells were collected by centrifugation at 2039 x g to be used for RNA extraction.

3.2.4.2 Suspension cultures

To maintain the suspension cultures, medium was aspirated into a 15 mL tube which was then centrifuged for 5 minutes at 2000 x g. The supernatant was removed and discarded and the pellet re-supended in fresh culture medium and the suspension transferred to a 25 cm^2 flask (T25). The flask was then incubated in a humidified incubator at 37 ºC with 5 % CO_2 for 24 hours after which the medium was checked for contamination visually. To harvest the non-adherent

culture for RNA extraction, the pellet would be re-suspended in PBS instead of culture medium.

3.3 miRNA extraction

Extraction of miRNA proved to be difficult in the study. Thus, a method to extract total RNA and synthesize miRNA cDNA using specific miRNA primers was devised.

3.3.1. Extraction of total RNA

The procedure for the extraction of miRNA was followed for both adherent and suspension cultures and was performed according to the manufacturers instructions (Macherey-Nagel NucleoSpin®). For the RNA extraction, the confluent cell lines were harvested as per section 3.2.2, at a concentration of 10^6 cells/mL. The cells were transferred to an Eppendorf tube and centrifuged at 2000 × g for 5 minutes at 4 °C. Thereafter, the supernatant was removed carefully avoiding disturbance of the pellet formed. The cells in the pellet were then lysed by adding 350 µL of lysis buffer RA1 containing 3.5 µL of a β-mercaptoethanol solution. The pellet was homogenised by vortexing vigorously for 10 seconds. The lysate was then transferred to a Nucleospin® Filter which was placed in a 2 mL collection tube. This was centrifuged for 1 minute at 11,000 x g using an Eppendorf 5417R bench top centrifuge. After centrifugation, the NucleoSpin® Filter was discarded and the RNA binding conditions of the lysate in the collection tube were adjusted with 350 µL 70 % ethanol. The solution was mixed

by pipetting up and down five times. The lysate was then pipetted onto a NucleoSpin® RNA Filter Column placed in a collection tube. This was then centrifuged for 30 seconds at 11, 000 x g. The column was transferred into a new 2 mL collection tube. The silica membrane of the column was desalted by adding 350 µL desalting buffer (MDB) to the column centrifuging for 1 minute at 11, 000 x g. The collection tube was discarded and the column placed in a clean tube. DNA present in the column was digested with 95µL of DNase (table 3). The column was kept at room temperature for 15 minutes and then washed with 200µL RA2 solution centrifuging at 11, 000 x g for 30 seconds. The flow-through was discarded and the column was placed back into the collection tube. A second and third wash were performed with 600 µL and 250 µL of RA3 solution respectively centrifuging at 11,000 x g for both washes. The collection tube was discarded and the column place in a new sterile tube into which RNA was eluted with 60 µL RNase free water. The concentration and quality of RNA was assessed using the Nanodrop ND-1000 spectrometer (ThermoScientific) and all the RNA samples were stored at -20 ºC.

3.4 Reverse transcription of miRNA to cDNA using stemloop sequence specific primers

Primers for the 5 miRNAs were designed by Khan, 2015 using the cotton estate Database available at http://www.leonxie.com/miRNAprimerDesigner.php

The cDNA was synthesized using the Transcriptor First Strand cDNA synthesis kit from Roche Life Sciences, according to the manufacturer's instructions. All

the reagents were kept on ice. The template RNA mixture was prepared with the reagents as shown in table 3.2 in a sterile, nuclease-free, thin walled PCR tube to a final volume of 13 µL. The tube was then incubated at 65 ºC for 10 minutes, after which the cDNA synthesis reagents in table 3.3 were added to make a final volume of 20 µL.

Table 3.2: Reagents for template RNA mix.

Reagent	Final concentration
RNA	1µg
Cocktail of stemloop sequence specific primers	2.5µM
PCR grade water	To make 13µL

As total RNA was extracted instead of miRNA, a cocktail of stemloop sequence specific primers was prepared to be used in cDNA synthesis. The cocktail was made by pipetting 1 µL of the reverse of each of the 5 primers into a nuclease-free, thin walled PCR tube. The concentration of each primer was 2.5 µM.

Table 3.3: Reagents for cDNA synthesis

Component	Volume	Final Concentration
Transcriptor Reverse Transcriptase Reaction Buffer	4 µL	1x (8mM $MgCl_2$)
Protector RNase Inhibitor	0.5 µL	20U
Deoxynucleotide Mix	2 µL	1 mM
Transcriptor Reverse Transcriptase	0.5 µL	10U
Final volume	20 µL	

The reaction was incubated at 55 °C for 30 min followed by a final inactivation step of 5 min incubated at 85 °C. The concentration of the synthesised cDNA was determined with a Nanodrop ND1000 Spectrophotometer.

3.5 Analysis of gene expression profiles of the miRNAs in cancer and control cell lines using qRT-PCR

Expression profiles of the five miRNAs were analysed via quantitative real-time PCR (qRT-PCR). The housekeeping miRNA miR-191a as well as the house

keeping gene GAPDH, were used as references. All reactions were performed on the LightCycler® 480 System (Roche Applied Science) instrument. The reactions were prepared as outlined in table 3.4.

Table 3.4: Reagents for a standard qRT-PCR reaction

Reagents	Final Concentration
SYBR Green Master Mix (10X)	1X
Forward Primer	1 µM
Reverse Primer	1 µM
cDNA	250ng
PCR Grade dH_2O	Variable to make 20µL
Final Volume	20 µL

A polymerase chain reacction was performed on each of the 5 miRNAs in each cell line. In addition, reactions for the reference miRNA primer 191a, reference housekeeping gene primer GAPDH and a no-template control (water) were also set up for each cell line. An aliquot of reaction mastermix was pipetted into each well of a 96 well plate and an aliquot of cDNA from each cancer cell line was then added as the PCR template to each well respectively. The experiment was set up with decreasing cDNA concentrations starting with 250 ng of cDNA to 0.0025 ng. Thus, 250 ng of cDNA for each cell line was run in duplicate for each primer

in order to construct a standard curve. A negative control was set up for each run containing 1 µL of PCR-grade water as a substitute for cDNA. The 96 well plates were sealed with clear sealing film and a qRT-PCR run set up on the LightCycler® 480 instrument according to the parameters in table 3.5. The evaluating parameters selected for data analysis were fluorescence (d[F1]/dT), melting temperature (Tm) and crossing point (Cp). The Second Derivative Maximum algorithm was employed for Cp determination where Cp was measured at the maximum increase of fluorescence.

Specificity of real-time PCR primers was determined by amplification plots, melting temperature, and melting curve analysis using LightCycler Software, Version 1.5 (Roche Diagnostics). Standard curves were generated using a dilution series in the concentration range 250ng to 0.0025 ng. The PCR efficiencies were calculated using the REST® software and all threshold cycle (Ct) values were taken into consideration according to the following equation: E=10[-1/slope] (Pfaffl 2002).

Table 3.5: Cycling Protocol for the qRT-PCR

Detection Format	**Block Type**	**Reaction volume**	
SYBR® Green	96 well	20µL	
Programme Name	Cycles	Analysis Mode	
Pre-incubation	1	None	
Amplification	40	Quantification	
Melting Curve	1	Melting Curve	
Cooling	1	None	
Programme Name	**Target (ºC)**	**Acquisition Mode**	**Hold (hh:mm:ss)**
Pre-incubation	95	None	00:03:00
Amplification	95	None	00:00:10
	Primer Dependent (65ºC)	None	00:00:20
	72	Single	00:00:10
Melting curve	95	None	00:00:05
	65	None	00:01:00
	97	Continuous	5-10 acquisitions/ºC

Cooling	40	None	00:00:10

3.6 Results and Discussion

This chapter aimed to investigate the utility of a panel of miRNAs (miR1, miR2, miR3, miR4 and miR5) predicted (*in silico*) to be dys-regulated in prostate cancer as potential diagnostic biomarkers. The main objective was to evaluate the expression of these miRNAs in a prostate cancer cell line, non-cancerous prostate cell line, benign prostate hyperplasia as well as a diverse cohort of other cell lines (table 3.1)

3.6.1 Normalisation and Statistical Analysis

The accuracy of qRT-PCR is heavily dependent on the proper normalization of expression data. There are several variables in a qRT-PCR experiment that need to be controlled for, both technical as well as biological variables. Technical variables include differences in sample collection, RNA extraction and target quantification and biological variables can be sample-to-sample inconsistency (Deo *et al*., 2011). Therefore, normalization is performed with the purpose to remove experimentally induced variation and to differentiate true biological changes. An inappropriate normalization of qRT-PCR data can lead to misleading conclusions (Peltier and Latham, 2008, Roberts *et al*, 2014). Thus, the choice of normalization method is a crucial step in data analysis.

miRNAs pose a significant challenge for normalization (Deo *et al*., 2011). This is thought to be due to the fact that miRNAs only represent 0.01 % of total RNA. Although this is a small fraction, it could have significant variation across different samples (Peltier and Latham, 2008). Despite these challenges, there are three normalization strategies which aid in expression profiling of miRNAs which include (i) average of all the quantification cycles values (Cq) from the experiments, (ii) stably expressed endogenous reference miRNAs, and (iii) external spike-in synthetic oligonucleotides. In this study, miR-191a (Schaefer *et al*, 2010; Deo *et al*, 2011), and GADPH (Ji *et al*, 2013) were used as reference nucleic acids for normalization, as they have been identified to be stably expressed in prostate tissue.

3.6.2 Standardization of qRT-PCR results

The miRNA expression studies were quantified using the Pfaffl model. This method requires the use of the Relative Expression Software Tool (REST®), a freely available Excel® based application that compares a sample group to a control group and calculates the relative expression between them (Pfaffl, 2004).

REST® uses a mathematical model (Pfaffl *et al*., 2002) that is dependent on the mean crossing point deviation between a sample and a control which are normalized by the mean crossing point deviation of a reference gene, in this case, a housekeeping miRNA.

The mathematical model in REST® relies on the determination of the crossing points (figure 3.1). The crossing point (Cp) or the threshold cycle (Ct) is the cycle

at which fluorescence achieves a defined threshold. It corresponds to the cycle at which a statistically significant increase in fluorescence is first detected in a qRT-PCR reaction (Heid *et al*., 1996; Rodriguez-Lazaro and Hernandez, 2013). This concept is the basis for accurate and reproducible quantification using qRT-PCR (Rodriguez-Lazaro and Hernandez, 2013). This is because a sample's Cp value depends on the initial concentration of DNA in the sample. A sample with a lower initial concentration of target DNA requires more amplification cycles to reach the Cp value. A sample with higher concentration requires fewer cycles. Thus, Cp values can be used as a quantitative measure of the input target number (Heid *et al*., 1996).

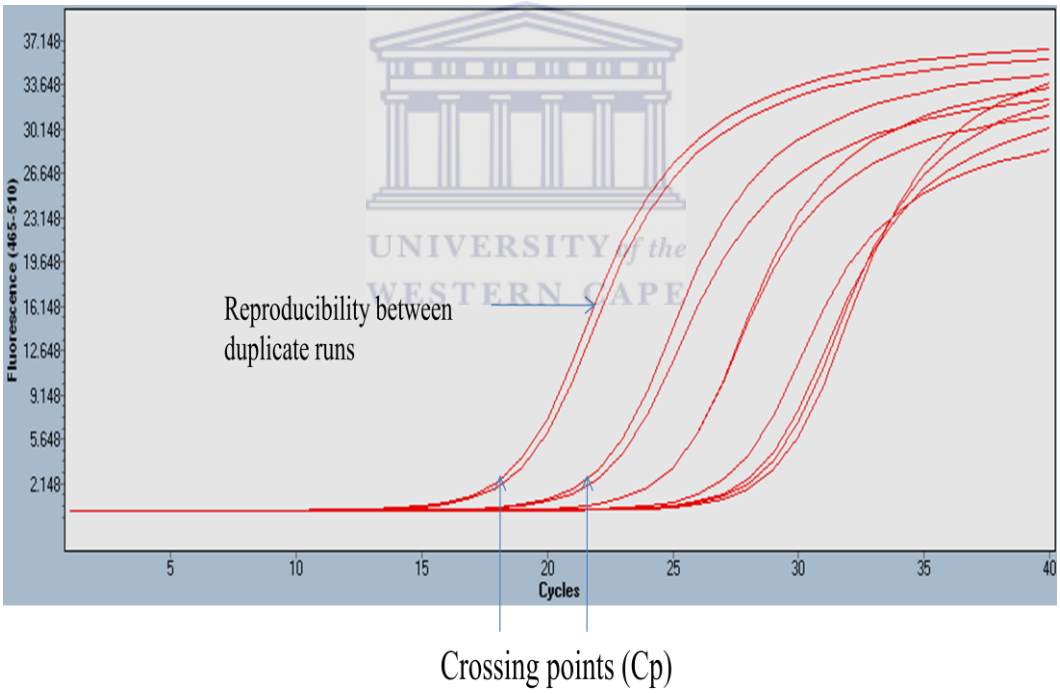


Figure 3.1: Amplification curve showing the crossing points at 19 cycles for one duplicate run and 21 cycles for another duplicate run. The amplification curve also shows reproducibility between replicates.

The REST® software aids in relative quantification analysis of qRT-PCR results. However, relative quantification is dependent on PCR efficiency (Pfaffl *et al*., 2002). The slope of the standard curve describes the kinetics of the PCR amplification. It indicates how quickly the amount of target nucleic acid (NA) can be expected to increase with the amplification cycles (Livak *et al*., 1997). The slope of the standard curve is also referred to as the efficiency of the amplification reaction. A perfect amplification reaction would produce a standard curve with an efficiency of 2, because the amount of target NA would double with each amplification cycle (Pfaffl *et al*., 2004). The PCR efficiency can easily be calculated using the formula:

$$E = 10^{-1/slope}$$

Equation 3.1: Efficiency of PCR is dependent on the gradient which describes the kinetics of the reaction. (Pfaffl *et al*., 2004).

An ideal slope would be -3.3 to give an efficiency of 2. However, a range of 3.2 to 3.6 is acceptable as the gradient (Livak *et al*., 1997; Pfaffl *et al*., 2004;Ruijter *et al*., 2009). Thus, a standard curve is a requirement in both relative and absolute quantification.

A relative quantification analysis was performed on the samples in this study. The expression levels of five miRNAs were evaluated in 11 cell line samples (table

3.1). A relative quantification analysis compares two ratios, the ratio of a target nucleic acid (NA) sequence to a reference NA sequence in an unknown sample and the ratio of the same two sequences in a standard sample, the calibrator (miR191a). This target sequence is the nucleic acid of interest, while the reference is a nucleic acid that is found at constant copy number in all samples and serves as endogenous control. The reference is used for normalization of sample-to-sample differences. The calibrator is typically a positive sample with a stable ratio of target-to- reference and is used to normalize all samples within one run (Roberts *et al*, 2014). Equation 3.2 shows how the result of this relationship is expressed.

$$\text{Normalized Ratio} = \left(\frac{\text{conc. Target}}{\text{conc. reference}} \right)_{\text{sample}} : \left(\frac{\text{conc. target}}{\text{conc. reference}} \right)_{\text{Calibrator}}$$

Equation 3.2: Normalization ratio for relative quantification analysis. (Pfaffl *et al*., 2004).

3.6.3 Importance of Melting Peak Analysis in qRT-PCR

The temperature at which a DNA strand separates or melts when heated can vary over a wide range, depending on the sequence, the length of the strand, and the GC content of the strand. For example, melting temperatures can vary for products of the same length but different GC/AT ratio, or for products with the

same length and GC content, but with a different GC distribution. Thus, a melting peak analysis can be performed to determine the characteristic melting temperature of the target DNA and to identify products based on their melting temperature (T_m). This distinguishes target amplicons from PCR artefacts such as primer dimers (Fraga *et al*., 2008). The KAPA SYBR FAST qRT-PCR kit optimized for LightCycler® 480 uses SYBR® Green I dye chemistry to detect the accumulation of an amplicon. SYBR® Green is a fluorogenic intercalating dye that emits a strong fluorescent signal upon binding to double-stranded DNA. In its unbound form, fluorescence is diminished (Nestorov *et al*., 2013).

3.6.4 Analysis and Quantification of qRT-PCR data

3.6.4.1 Analysis of qRT-PCR normalisation

Normalisation of qRT-PCR is important because of the many variables that need to be controlled for. Figure 3.2 shows the comparison of miR1 expression in KMST-6 and A549 before normalization with the housekeeping miRNA, miR191. The miRNA is down-regulated in both instances. However, the factor by which the miRNA is down-regulated is 8.626 before normalisation and 15.66 after normalization. The p-values are 0.0895 and 0.0435 respectively.

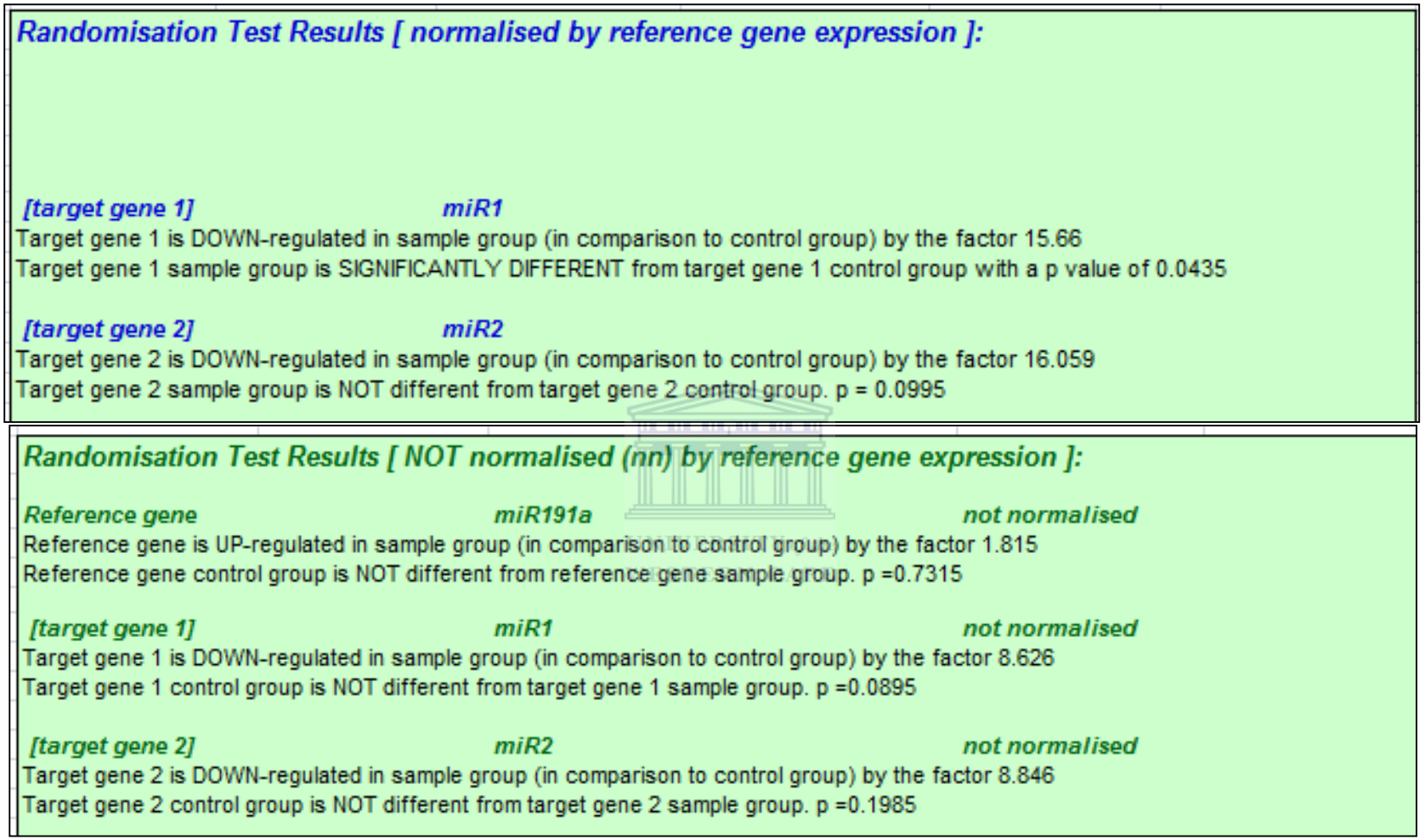


Figure 3.2: Comparison of miR1 expression between the KMST-6 cell line and A459 cell line, before the normalization test (green writing) and after performing normalization (in blue).

3.6.4.2 Analysis of amplification curves

Figure 3.3 shows the amplification curves of miR1 in KMST. The figure shows reproducibility of the duplicate runs.

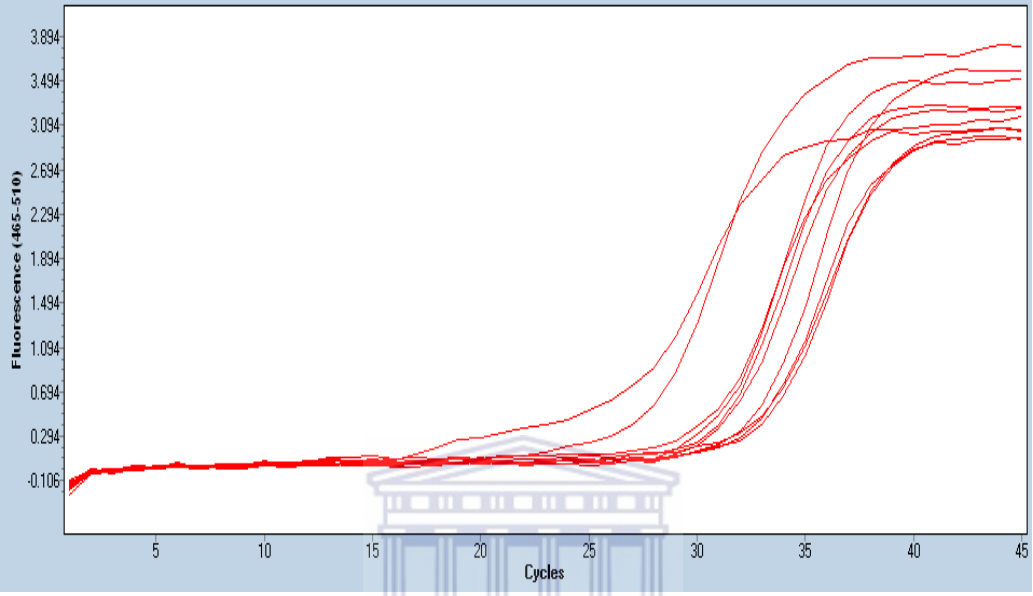


Figure 3.3: Amplification curve of miR1 in the KMST cell line. The different amplification curves represent different concentrations of KMST cDNA from 250 ng to 0.0025 ng in duplicate from left to right.

Examining figure 3.3, it can be seen that the reproducibility in the 250 ng duplicates was not consistent. This could have been caused by a number of reasons. One could be the introduction of random errors; this is indicated by the observation that one duplicate has a higher Cp value than the other duplicate. This could have resulted from adding more cDNA to the first duplicate hence demonstrating the importance of sample normalization as well as robust statistical analyses so that the results reflect an accurate biological event within each cell

line. Comparing the amplification curves of figure 3.3 and 3.4, it can be observed that the latter shows better evidence of varying cDNA concentration than the former.

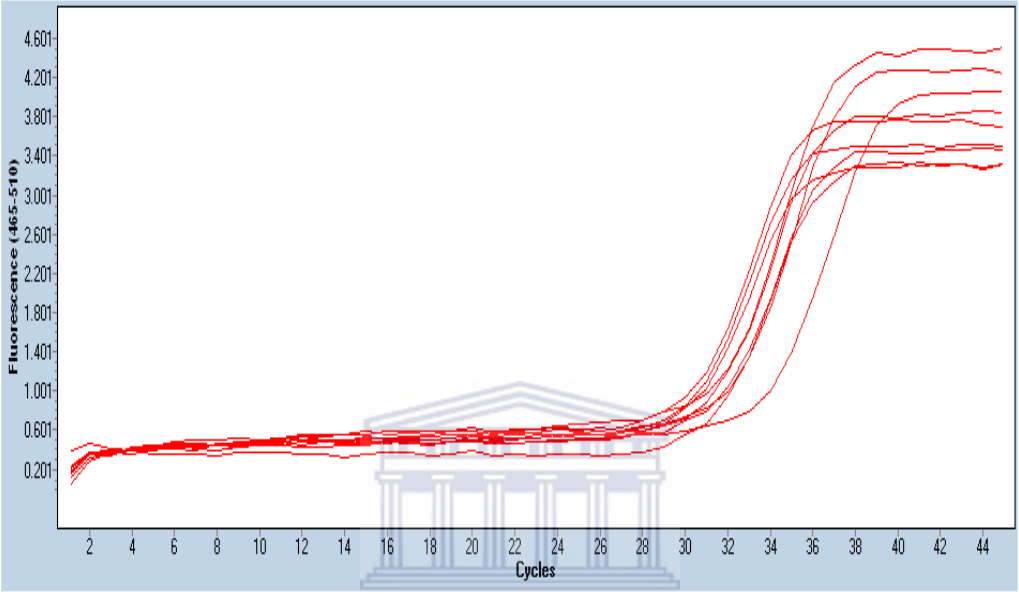


Figure 3.4: Amplification curve of miR1 in BPH1 cells. The different amplification curves represent different concentrations of BPH1 cDNA from 250 ng to 0.0025 ng in duplicate from left to right.

3.6.4.3 qRT-PCR melting peak analysis

A melting curve analysis is important to determine the homogeneity of the PCR product of concern. Figure 3.5 shows the melting peak of miR1 in BPH1 cells. A prominent peak is seen at the T_m of 83 ºC. This corresponds to the expected T_m of miR1 which was calculated from the miRNA sequence obtained from mirBase. Figure 3.6 shows the melting peak of the housekeeping miRNA in the BPH1 cell

line. Examining figure 3.5 shows that there are 3 more additional peaks at 67 ºC, 79 ºC and 91 ºC in addition to the expected peak at 83 ºC.

One explanation of this is that DNA melting curves depend on the G/C content of a DNA strand (Rodriguez-Lazaro and Hernandez, 2013). As the dsDNA starts to melt, regions of the amplicon that are more stable (i.e G/C rich) do not melt immediately (Draghici *et al*., 2008; Rodriguez-Lazaro and Hernandez, 2013). These stable regions maintain their dsDNA configuration until the temperature is sufficiently high to cause it to melt. This scenario results in 2 melting phases (Draghici *et al*., 2008; Nestorov *et al*., 2013; Rodriguez-Lazaro and Hernandez, 2013). The data in figure 3.5 is consistent with this interpretation. Additional sequence factors can also cause products to melt in multiple phases. These include amplicon misalignment in A/T rich regions, and designs that have secondary structure in the amplicon region (Rodriguez-Lazaro and Hernandez, 2013). Melting curves for the rest of the miRNAs in the various cell lines can be found in appendix B.

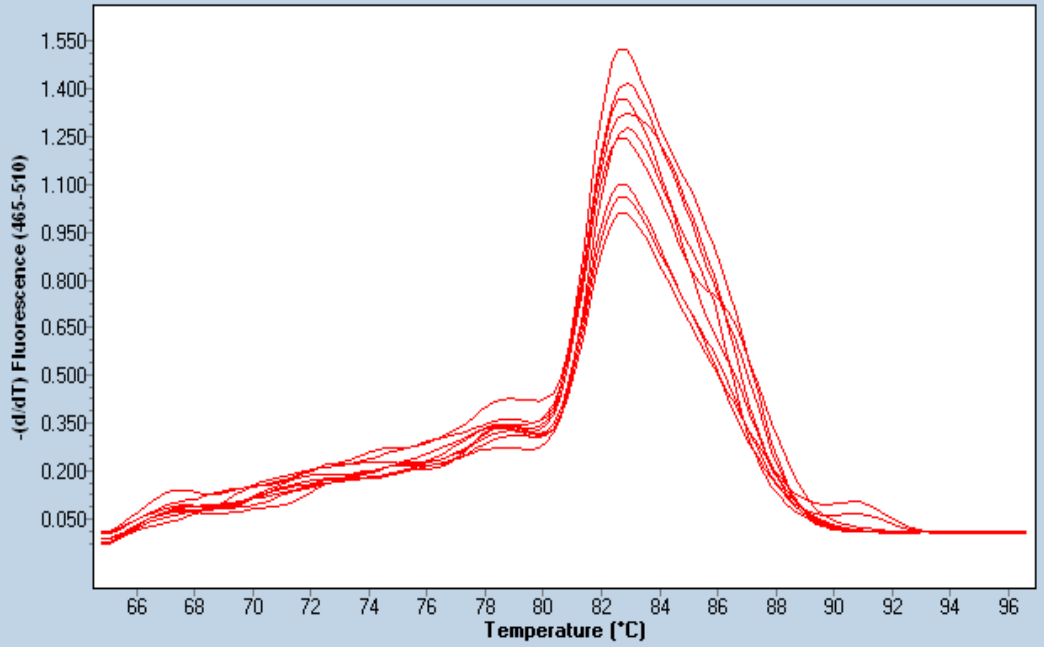


Figure 3.5: Melting peak of miR1 in BPH1 cells. A prominent peak is seen at the T_m of 83 ºC.

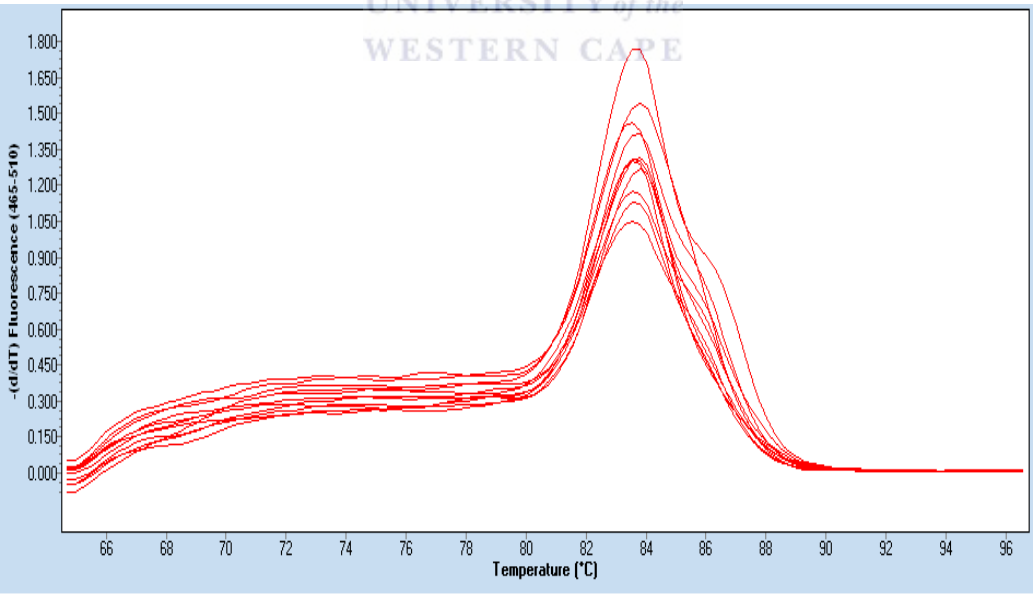


Figure 3.6: Melting peak of miR191 in BPH1 cells. A prominent peak is seen at the T_m of 84 ºC.

3.6.4.4 Analysis of qRT-PCR Gene Expression data

An analysis was done to evaluate the specificity of the putative miRNA as prostate cancer biomarkers. This was done by analysing their expression patterns in benign prostate hyperplasia, BPH1, LNCaP, PNT1a as well as in a panel of six different cancer cell lines (refer to table 3.1). The KMST cell line, a normal skin fibroblast was used for sample normalisation.

The REST® software calculates relative expression between two samples using a statistical model called the Pair Wise Fixed Reallocation Randomization Test to normalise data. The data obtained were imported into an Excel® spreadsheet to create a graph showing the relative expression levels of miR1. miR2, miR3, miR4 and miR5 in the various cancer cell lines. This expression was relative to the expression levels of the miRNAs in the control samples, where the control samples were given an arbitrary value of one which indicates no variation of regulation of the miRNAs of interest.

Differential expression of the 5 miRNAs was observed across all cancer cell lines (figure 3.7), with miR2, miR3 and miR5 being significantly highly differentially expressed in the prostate cancer cell line LNCaP when compared to the other cancer cell lines. The expression ratios were 8.79, 13.87 and 15.32 respectively (Table 3.6). Thus, these miRNAs have the highest potential to be biomarkers for prostate cancer. Also observed, was the expression of miR1 in the benign prostate hyperplasia cell line (BPH1). The miRNA is significantly highly expressed in BPH1 when compared to other cancer cell lines, (figure 3.4), with an expression ratio of 4.85 (table 3.6). Currently, there are no molecular markers

identified for clinical use for detection of benign prostate hyperplasia or its likelihood to progress to cancer (Cannon and Getzenberg, 2012; Kunar *et al*., 2013).

DIANA-Tarbase generated pathways indicated that miR1 may be involved in various cancer causing related pathways including regulation of actin cytoskeleton and glucosaminoglycan biosynthesis (figure 2.6). Studies have shown that genes involved in actin cytoskeleton regulation, calcium signalling and glucosaminoglycan biosynthesis are targeted in BPH (Savli *et al*., 2008; Endo *et al*., 2009). Thus, this could explain the predicted association of miR1 in these pathways. However, more studies would have to be conducted on miR1 to determine its potential as a good candidate for benign prostate hyperplasia.

Examining miR2, it was observed that it was significantly highly expressed in LNCaP when compared to the other cell lines. Expression of miR2 was down-regulated in BPH1, A459, HEPG2, HeLa and H157 by ratios of -2.4, -4.0, -4.10, -7.82 and -4.28 respectively (table 3.6). The miRNA was up-regulated in MCF7 by a ratio of 0.85 in HT29 by 1.09 and in LNCaP by 8.79. Thus, miR2 is could be considered as a possible good indicator of prostate cancer. From DAVID, miR2 was predicted to target the gene AFAP1 (Table 2.5). This gene is known to encode a protein that is a potential modulator of actin filament integrity which responds to cellular signals (Garzon *et al*., 2014). This gene is not well characterized in literature. However, further prediction studies in STRING indicated that the gene is involved in protein cancer networks and it is linked to another gene CTNND1 which is regulated by miR4.

GeneHub-GEPIS showed that AFAP1 is expressed in both normal and tumour prostate tissues (figure 2.16). However, the expression of this gene in the tumour tissue was shown to be down-regulated. This could be as a result of the action of over-expression of miR2. Further studies would have to be undertaken to validate this.

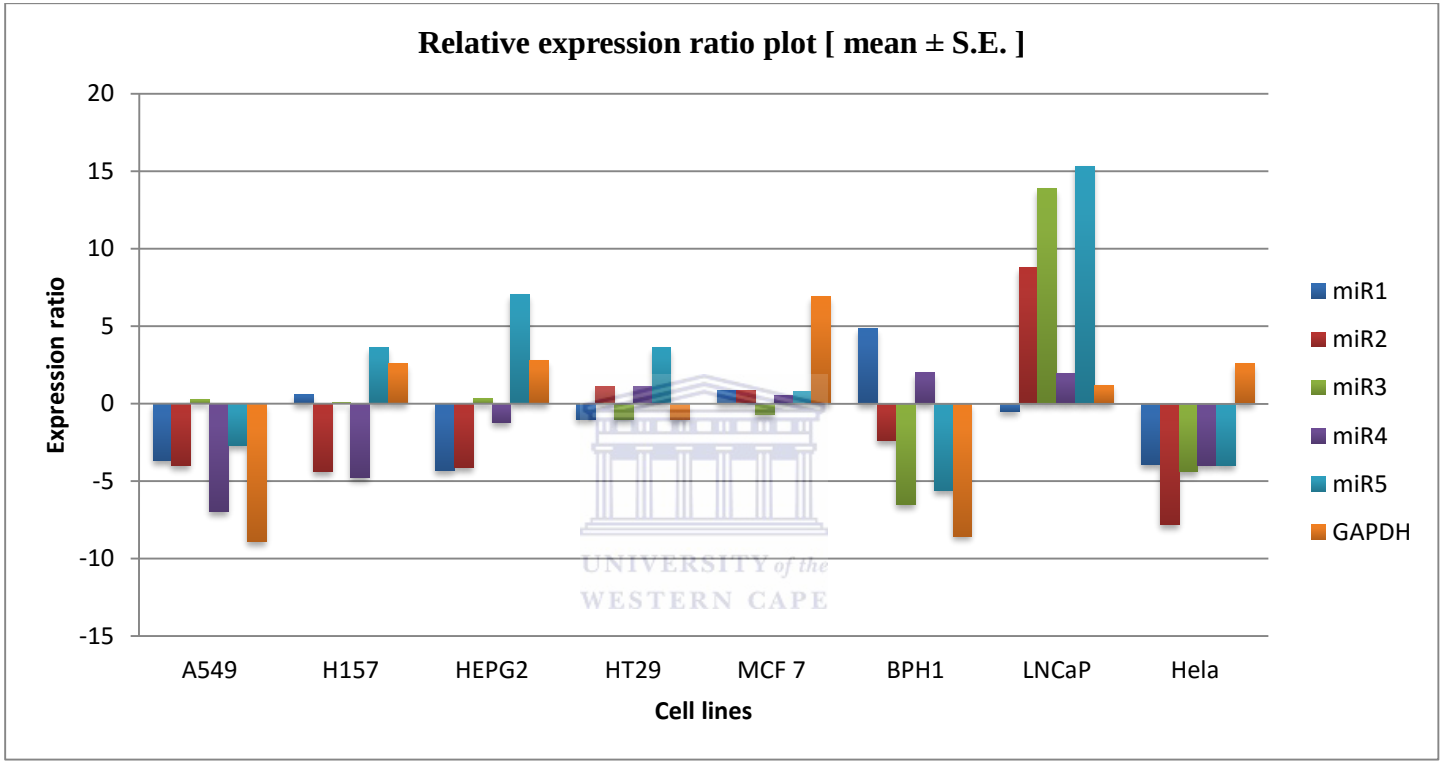


Figure 3.7: Relative expression ratio plot of the five miRNAs in various cancer cell lines including prostate cancer (LNCaP) and benign prostate hyperplasia (BPH1). The bars indicate up-regulation (above 0) of the miRNAs and down-regulation (below 0) of the miRNA.

Table 3.6: Fold expression ratios of the five miRNAs

	BPH1	LNCaP	A549	HEPG2	MCF7	HeLa	HT29	H157
miR1	4.85	-0.49	-3.96	-4.32	0.82	-3.94	-1.40	0.6
miR2	-2.4	8.79	-4.0	-4.10	0.85	-7.82	1.13	-4.38
miR3	-6.51	13.87	0.3	0.34	-0.69	-4.38	-1.04	0.11
miR4	1.99	1.92	-6.99	-1.2	0.50	-4.01	1.11	-4.79
miR5	-5.61	15.34	-2.69	7.057	0.79	-4.00	3.59	3.59
GAPDH	-8.560	1.189	-8.92	2.78	6.91	2.58	-1.01	2.58

An examination of the expression of miR3 shows that it is significantly over-expressed by a ratio of 13.87 in the prostate cancer cell line LNCaP (table 3.6). Table 3.6 also shows that miR3 is under-expressed in BPH1 by a ratio of -6.51. Comparing the expression ratios of miR3 in BPH1 and LNCaP indicates that it could serve as a good indicator of prostate cancer. The under-expression of miR3 in BPH1 could help in the specificity of prostate cancer diagnosis. This is because the current most widely used biomarker PSA cannot specifically distinguish between benign prostate hyperplasia and prostate cancer (Velonas *et al*., 2013; Qu *et al*., 2014). PSA is organ-specific and not disease-specific. Thus, it is prone to high false-positive diagnosis (Heidenreich *et al*., 2011; Velonas *et al*., 2013; Qu *et al*., 2014) as there are several reasons, in addition to prostate cancer, for elevated levels of PSA found in a man's blood.

From the *in silico* study, miR3 was predicted to regulate two genes, MNT and LIG4. The MAX binding protein, (MNT) has been found to be highly expressed in prostate tumour tissue with a digital expression unit (DEU) of 94.6 (figure 2.8). However, the gene LIG4 is not expressed in prostate tumours. The qRT-PCR analysis showed that miR5 was the most over expressed miRNA with an expression ratio of 15.32 in LNCaP (table 3.6). It was not highly expressed in BPH1, A549 and HeLa with expression ratios of -5.61, -2.6 and -4.0 respectively. It was significantly highly expressed in HEPG2 with an expression ratio of 7.057. Expression levels in H157, HT29 and MCF7 were 3.59, 1.61 and 0.50 respectively. Thus, it could be a good indicator of prostate cancer. However, its level of expression in HEPG2 (liver cancer) takes away from its specificity to prostate cancer. The miRNA was predicted to be involved in the regulation of the

genes YWHAZ and TNFSF13B (table 2.5). Several studies have demonstrated that YWHAZ has a pivotal role in tumour cell proliferation through its over-expression (Chen *et al*., 2012; Nishimura *et al*., 2013). In many cancers, TNFSF13B has been shown to be an endogenous inhibitor of neovascularisation. This is a critical component of the negative control mechanism that operates in normal cells but is missing in tumour cells (Vassiliki *et al*., 2008; Deng *et al*., 2012).

3.6.5 Summary and conclusion

The validation of biomarkers is a critical step in the biomarker discovery pipeline. In order to become a clinically approved marker, a potential biomarker should be confirmed and validated using hundreds of specimens. The tests should also be reproducible, specific and sensitive (Drucker and Krapfenbauer, 2013). Microarray technologies have previously been used to identify differentially expressed genes and have been used in numerous studies pertaining to human malignancies. However, microarray results have been influenced by various sources of variability including minor changes in experimental conditions such as biological heterogeneity in the population as well as in the specimen, specimen collection and handling and RNA extraction and amplification (Murphy, 2002; Abdulla-Sayani *et al*., 2006; Draghici *et al*., 2008). This would make it difficult to reproduce the results. As a consequence, differentially expressed genes in such preliminary discoveries could be confirmed using alternative methods such as qRT-PCR (Hu *et al*., 2006). Quantitative real time PCR quantifies small changes

in gene expression and thus can give insight into the role of a gene and/or its product. These changes in gene expression can be indicative of a diseased state as the body tries to maintain homeostasis (Pfaffl, 2001).

In this study, we evaluated the expression profiles of five miRNAs that were predicted via *in silico* methods to play regulatory roles in cancer onset and progression. Examination of the expression of these miRNAs in a panel of cancer cell lines showed that miR1 is over-expressed in BPH1. Thus it has the potential to serve as an indicator of benign prostate hyperplasia. However, because of its expression in MCF7 and H157 albeit minutely (0.8 and 0.6 respectively); it may have to be used in conjunction with other biomarkers or processes. It was also clear that miR3 was over-expressed in LNCaP and not in any other cell line. The expression ratio of miR3 in LNCaP was found to be 13.87. This makes it a good biomarker candidate for prostate cancer diagnosis. The study also showed that miR2 could be another potential biomarker for prostate cancer diagnosis. It had an expression ratio of 8.79 in LNCaP. However, it is also expressed in two other cancer cell lines HT29 and MCF7 with expression ratios of 1.09 and 0.85 respectively. However, one miRNA may not be sufficient for the detection of a condition (Carlsson *et al.*, 2011). Thus, both miRNAs could be used in combination with other biomarkers. Additionally, more cell lines and patient samples would need to be evaluated to establish the specificity of the expression of these miRNAs in BPH1 and prostate cancer. The use of a singular miRNA, or a combination of miRNAs could potentially improve the predictive accuracy or prognostics as well as treatment outcomes in prostate cancer (PCa). The use of these markers could play an important role in screening for PCa. This is because

current methods are invasive, and painful. Identifying biomolecules that can be present in bodily samples like urine could reduce unnecessary biopsies.

Chapter 4

4.1 General discussion and future work

Prostate cancer (PCa) is the most frequent tumour in men and a major cause of cancer-related morbidity and mortality (WHO, 2015). According to the International Agency for Research on Cancer (GLOBOCAN) in 2012, prostate cancer was among the most commonly diagnosed cancers in males, coming second after lung cancer. Additionally, prostate cancer makes up 8% of all cancers diagnosed in the world. In 2013, 238,590 men were diagnosed with cancer of the prostate and 29,720 men died as a result. Approximately 4500 of the deaths related to PCa were in South Africa (Cancer Association of South Africa, 2014) which makes PCa a global epidemic. Current diagnostic tools include digital rectal examinations (DRE) (Schröder *et al.,* 1998), prostate specific antigen test (PSA) (Lu-Yao et al., 2003), biopsy (Essink-Bot *et al.,* 1998) and ultra sound (Bonekamp *et al.,* 2011). All these diagnostic methods currently being employed are however, invasive, lack specificity and sensitivity (ACS, 2014; Djulbegovic *et al.,* 2010). Additionally, the diagnostic application of the prostate specific antigen (PSA) has led to widespread over-diagnosis and subsequent overtreatment of clinically insignificant tumours (Simmons *et al.,* 2011; Stavridis *et al.,* 2010; Wolf *et al.,* 2010). It is foreseen that this problem may increase in the future (Bitu, 2015; Siegel, 2015). Therefore, there is a need for a less invasive early detection method with the ability to overcome the lack of specificity and sensitivity. Biomarkers have recently been identified as a viable option for early detection of disease for example biological indicators i.e. DNA, RNA, proteins and microRNAs (miRNA).

Mature miRNAs are a class of naturally occurring, small non-coding RNA molecules. They are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to down-regulate protein expression in a variety of manners, including translational repression, mRNA cleavage, and de-adenylation (Filipowicz, 2005; He *et al.,* 2005; Mraz *et al.,* 2009). miRNAs are becoming increasingly recognized as powerful biomarkers for human disease. The information potential held by miRNAs, combined with the fact that they are stable in serum and plasma, has led to a rapidly growing interest in using miRNAs in blood or urine as diagnostic and prognostic biomarkers (He *et al.,* 2005). The aim of this study was therefore to identify and characterize miRNAs, as a class of less invasive biomarkers for early diagnosis of prostate cancer.

A number of miRNAs have been shown to influence key cellular processes involved in prostate tumourigenesis, including negative regulation of apoptosis, cell proliferation and migration and the androgen signalling pathway (Carlsson *et al*., 2011). Currently, a few studies have been undertaken to identify miRNA specific for prostate cancer Ambs *et al*., 2008 suggested that miRNA expression alters the development and progression of prostate cancer and some of the cancer-related genes are regulated by miRNAs (Ambs *et al*., 2008). Porkka *et al*., also identified 51 miRNAs that are differentially expressed between benign and malignant prostate tumours, of which 37 were down-regulated and 14 up-regulated (Porkka et al., 2007). These differentially expressed miRNAs lead to alteration in the expression and activity of their targets in prostate cancer. MiR-21 in prostate cancer has been shown to stimulate androgen-dependent cell growth

and can also rescue cells from androgen deficient growth arrest. This indicated that miR-21 may also mediate castrate-resistant prostate cancer (CRPC) development (Jackson *et al*., 2014). Several studies, (Li *et al*., 2012; Brase *et al*., 2011; Jackson *et al*., 2014) recently established that tissue miR-21 expression levels may have clinical importance. They evaluated miR-21 expression levels in a cohort of 169 radical prostatectomy tissue samples and found that increase in miR-21 expression levels was associated with pathological stage of prostate cancer. They went on to demonstrate *in vivo* tumour growth repression in a mouse model treated with a miR-21 inhibitor.

Brase *et al*., 2011 also examined miR-141 as an miRNA correlated with tumour progression in cancer. The study identified it as one of the miRNAs present at high levels in patients with malignant tumours. In the same study, MiR-141 was also shown to distinguish between patients with metastatic prostate cancer and healthy controls. However, both miR21 and miR141 are implicated in other cancers apart from prostate cancer and thus may not be suitable diagnostic markers for the disease but rather have prognostic value, (Li *et al*., 2012; Brase *et al*., 2011; Jackson *et al*., 2014). Thus, the need to identify putative novel miRNAs is still a necessity.

The study investigated a panel of five miRNAs implicated in prostate cancer as putative biomarkers for detection of the disease. The panel was mined from online databases and literature mining in a previous study (Khan, 2015). It is well known that miRNAs control the expression of genes involved in various biological processes (Bartel, 2009). This is done by binding to the 3' UTR of their

target mRNA (section 1.6). Thus, target gene prediction of miRNA is important in identifying which genes are affected by miRNA dys-regulation in diseases including prostate cancer.

Target prediction of miRNA target genes was done using TargetScanHuman and miRDB. These are platforms that are publicly available. The target genes identified for the five miRNAs via TargetScanHuman were 1502. miRDB predicted 1076 target genes for the miRNAs. It was interesting to see a 28 % difference in the results from the two databases. This could be attributed to the properties of the two databases. The combined list from miRDB and TargetScanHuman were used as input into DAVID. This tool allows for functional annotation of genes. Thus, the processes that the genes are involved such as cellular component, biological function and molecular function are discerned.

The list of genes in DAVID were reduced by looking at the enrichment score. An enrichment score greater than 1.3 was considered significant. The enrichment score is a figure of how important the particular gene is in the submitted gene list. Literature mining of the genes was also performed. Those genes involved in PCa but not experimentally validated were prioritized.

It was found that terms such as regulation of apoptosis (figure 2.3) in biological processes are crucial to cancer onset. In the same way, genes with molecular functions such as metal ion binding and nucleoside binding were prioritized

(figure 2.4). As were genes with their cellular components localised to the cell surface (figure 2.2). Table 2.5 summarizes the miRNAs and the genes they target. Analysis of gene interaction networks in STRING showed that the genes are involved in protein networks with each other as well as with genes such as RAF1 which is a proto-oncogene (Mikula *et al*., 2001) and CASP3 which is responsible for apoptosis execution (Jin *et al* ., 2007). DIANA-Tarbase generated a heat map that indicated the involvement of the five miRNAs in cancer causing pathways, such as transcription mis-regulation in cancer as well as lipid metabolism. The miRNAs involved in these pathways included miR1, miR3 and miR5.

Even though *in silico* methods are a cost effective, easy way of identifying novel biomarkers and maybe a tool of choice by many scientists, they generate a plethora of data that should be analysed critically. It is also important to validate *in silico* results. Thus, a molecular approach was employed to compare expression of the 5 miRNAs predicted to be involved in gene regulation during prostate cancer.

In this study, we evaluated the expression profiles of the five miRNAs that were predicted via *in silico* methods to play regulatory roles in prostate cancer. We investigated expression profiles of the five miRNAs using qRT-PCR. This was done in various cell lines including the PCa cell line, benign hyperplasia cell line, normal prostate cell line, KMST, MCF12 and a panel of 6 other cancer cell lines. The expression profiles of the miRNAs in the cell were analysed using the relative expression tool REST®.

A hypothesis was made that the miRNAs would be significantly highly expressed in LNCaP and that additionally, some of the miRNAs may also show significant high expression in the BPH1. It was observed, as discussed in 3.6 that miR1 was significantly expressed in BPH1 and miR3 was highly expressed in LNCaP. Thus these hold potential to be biomarkers for BPH and PCa upon further work.

However, because of the expression of miR1 in MCF7 and H157 albeit minutely (0.8 and 0.6 respectively); It may have to be used in combination with another biomarker. The expression ratio of miR3 in LNCaP was found to be 13.87. This makes it a good biomarker for prostate cancer diagnosis. It has been reported that using more than one miRNA for cancer detection is more appropriate (Carlsson *et al*., 2011). Thus, additional cell lines and patient samples would need to be evaluated to establish the specificity of the expression of miR3 in prostate cancer. The study also showed that miR2 could be another potential biomarker for prostate cancer diagnosis. It had an expression ratio of 8.79 in LNCaP. However, it is also expressed in two other cancer cell lines HT29 and MCF7 with expression ratios of 1.09 and 0.85 respectively. Thus, it could be used in combination with another biomarker.

The use of a singular miRNA, or a combination of miRNAs could potentially improve the predictive accuracy or prognostics as well as treatment outcomes in prostate cancer (PCa). The use of these markers could play an important role in screening for PCa, however the objective remains to reduce the number of unnecessary biopsies performed and to limit the invasive procedures performed to differentiate between normal cells, benign tumours and PCa.

There is evidence from *in vitro* and *in vivo* studies that alteration in miRNA function plays a role in prostate carcinogenesis (Berger *et al.,* 2014 and Murata *et al.,* 2012). miRNA dysregulation influences a number of critical cellular processes involved in carcinogenesis, including but not limited to: stimulation of the cell cycle, avoidance of apoptosis, epithelial mesenchymal transition and modulation of AR-mediated signalling (Ozen *et al.,* 2008; Clape *et al.,* 2009; Fu *et al.,* 2010; Schaefer *et al.,* 2010).

Further understanding of the expression patterns of miRNA dys-regulation may allow the development of novel diagnostic and therapeutic strategies involving miRNA augmentation or inhibition in the future. The potential to detect circulating miRNAs in serum and potentially, in urine, clearly exists (Chen *et al.,* 2008; Mitchell *et al.,* 2008; Li *et al.,* 2007; Park *et al.,* 2009). The present study revealed miRNAs that may play a role in detection of prostate cancer and have demonstrated over- or under-expression when comparing non- cancerous prostate cells to prostate cancer cells. Further investigation of these miRNAs in larger patient groups will help to define their potential role as diagnostic and prognostic biomarkers in the future.

Future work would include examination of the expression profiles of the miRNAs in a larger panel of cells as well as investigation of expression profiles of these miRNAs in patient samples such as blood, urine and saliva. It would also be interesting to analyse the UTR sequences of the miRNAs targets experimentally. This will prove that the target genes identified using the *in silico* methods are indeed regulated by these miRNAs. Once the target genes are confirmed, further analyses can be performed on the miRNAs to study their roles

in prostate cancer development. This can be done by performing knockdown of the miRNA via Locked Nucleic Acid modified probes. This will aid in a wet bench functional analysis of the miRNA.

Appendix A

Chapter 2 supplementary information.

TiGER and GeneHub-GEPIS expression profiles of the of FOXC1, TNFSF13B, BIRC2, CTNND1, TNFSF15 and AFAP1 genes.

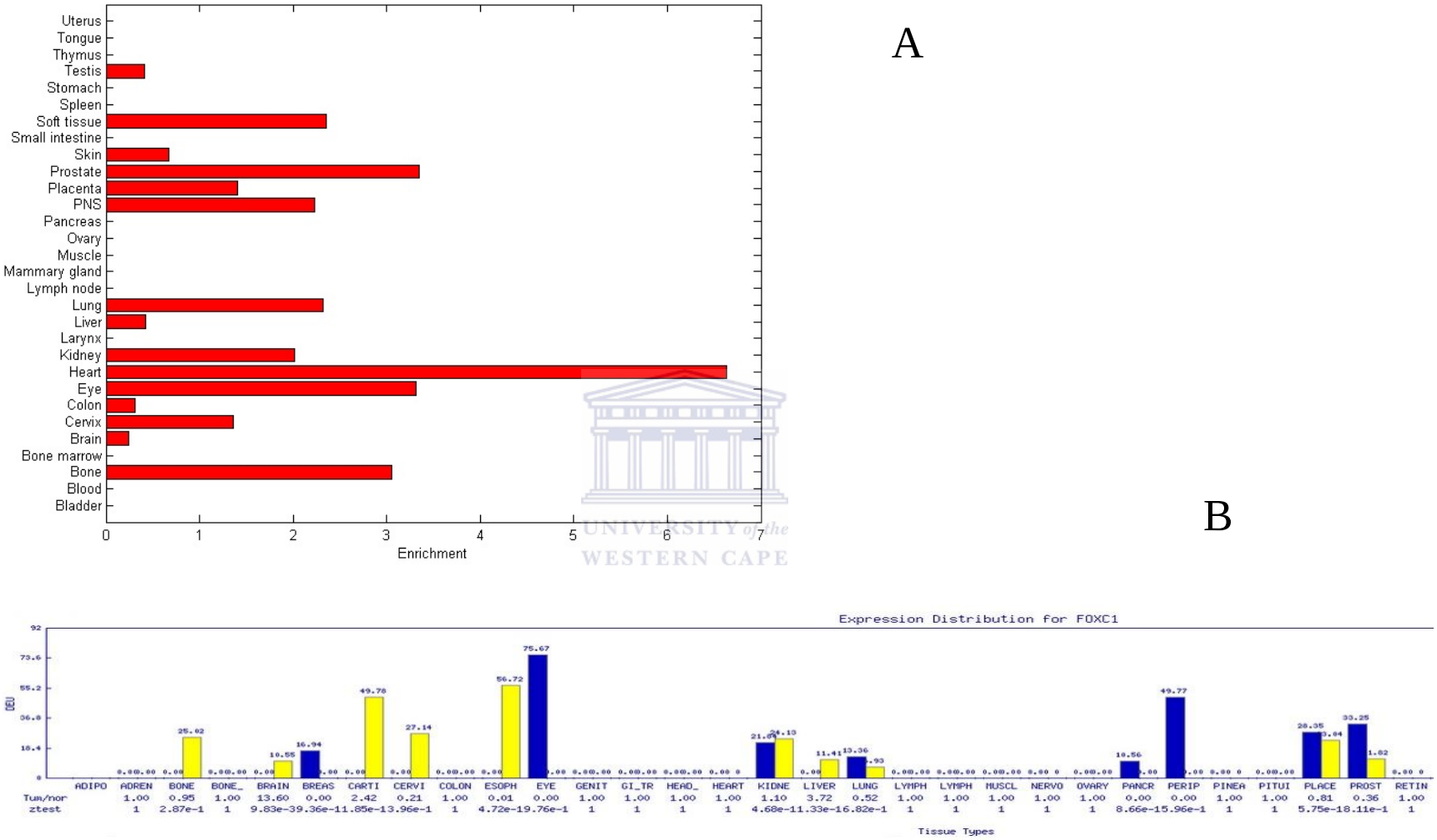


Figure 2.11: Expression profile for FOXC1 from TiGER (A) FOXC1 is preferentially expressed in the heart with an expression enrichment value greater than 6. Expression in prostate tissue is between 3 and 4. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

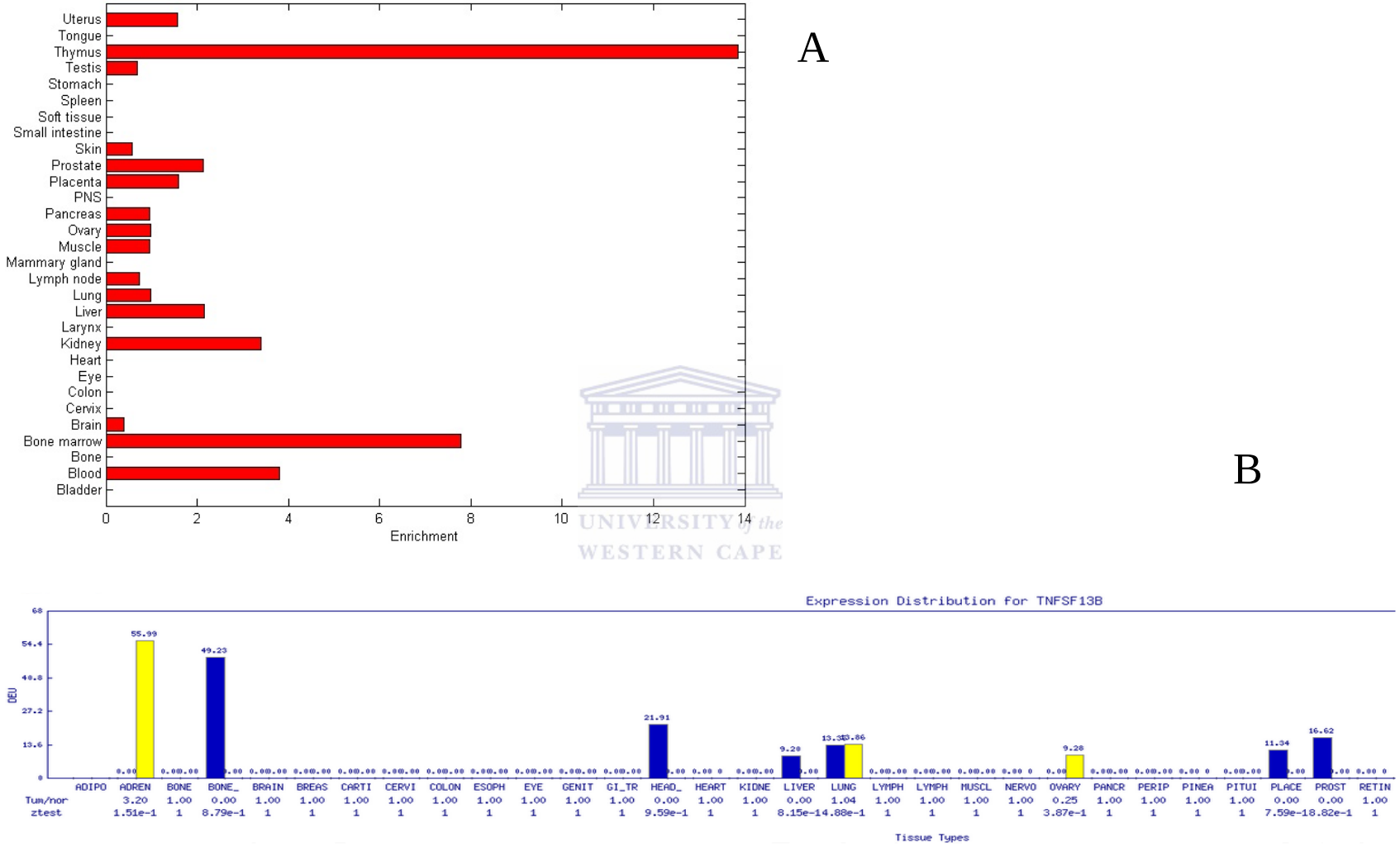


Figure 2.12: Expression profile for TNFSF13B from TiGER (A) TNFSF13B is preferentially expressed in the thymus with an expression enrichment value greater than 13. Expression in prostate tissue is about 2.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

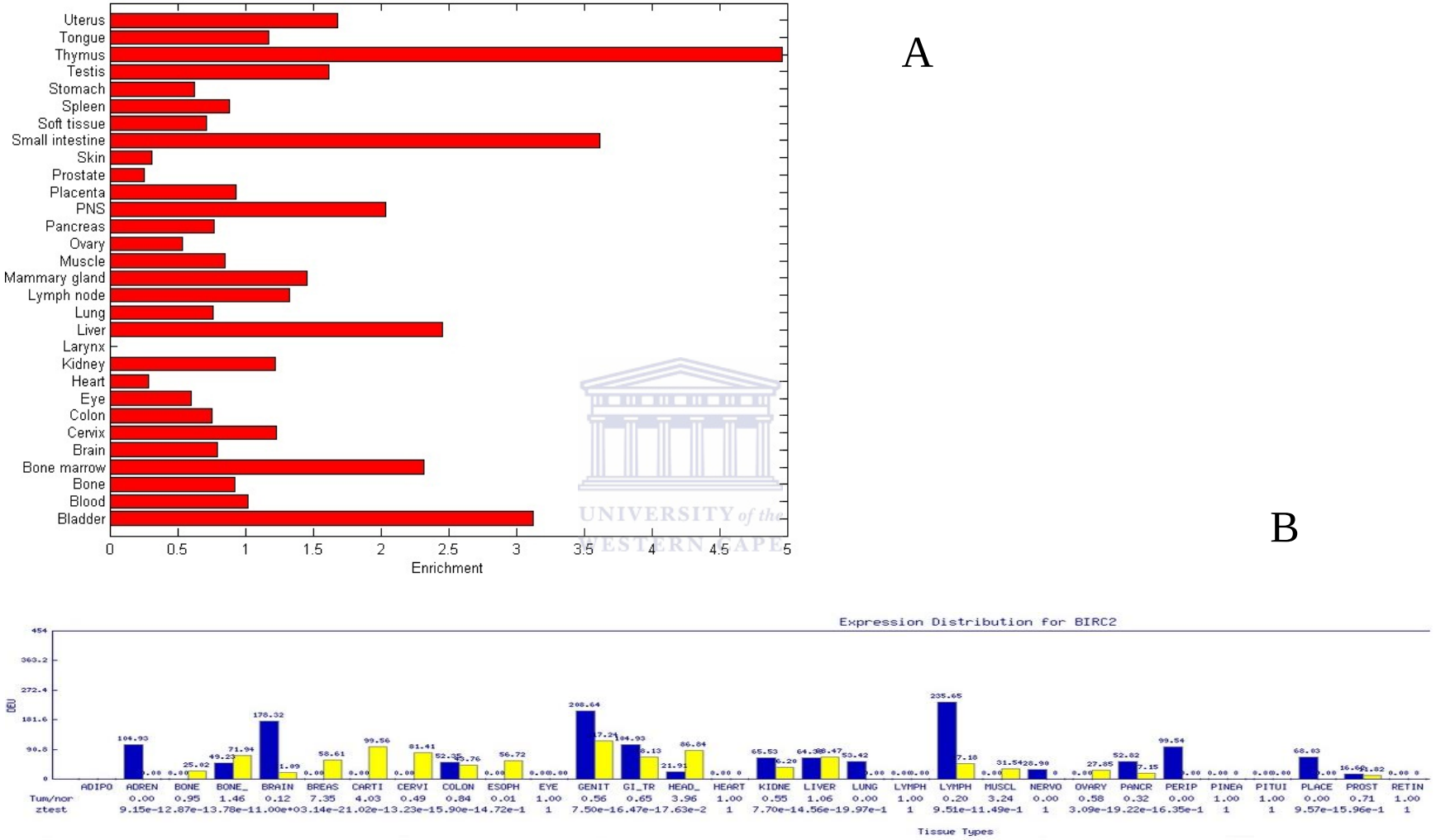
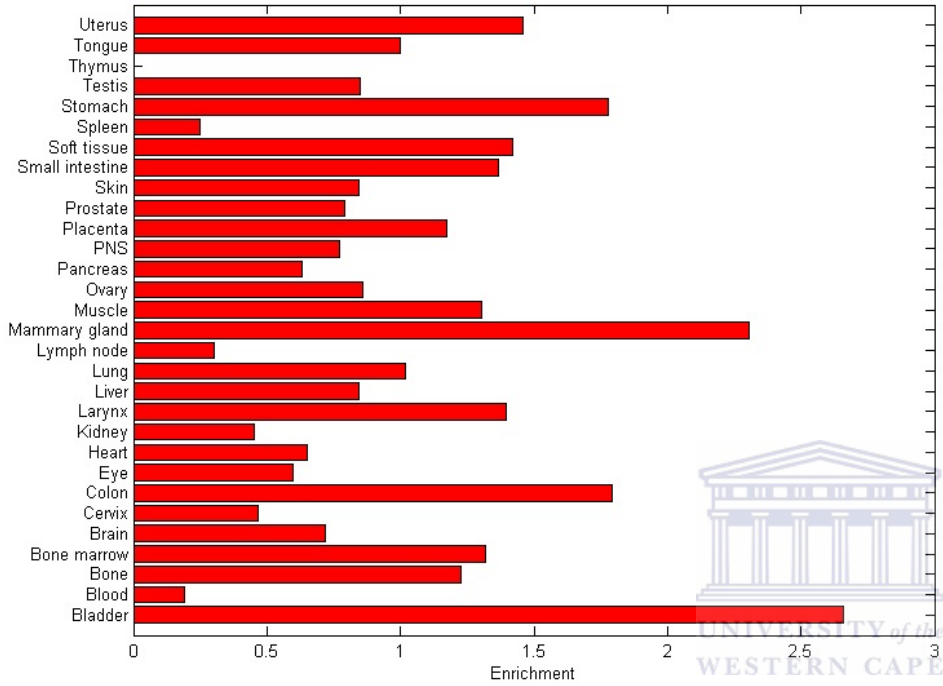


Figure 2.13: Expression profile for BIRC2 from TiGER (A) BIRC2 is preferentially expressed in the thymus with an expression enrichment value greater than 4.5. Expression in prostate tissue is less than 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.



A

B

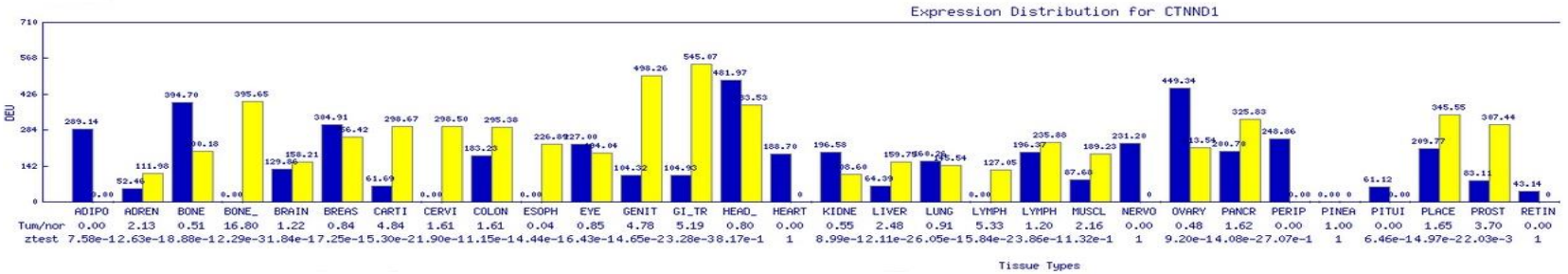


Figure 2.14: Expression profile for CTNND1 from TiGER (A) CTNND1 is preferentially expressed in the bladder with an expression enrichment value greater than 2.5. Expression in prostate tissue is between 1 and 0.5. **GeneHUB-GEPIS (B).** Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow.

B

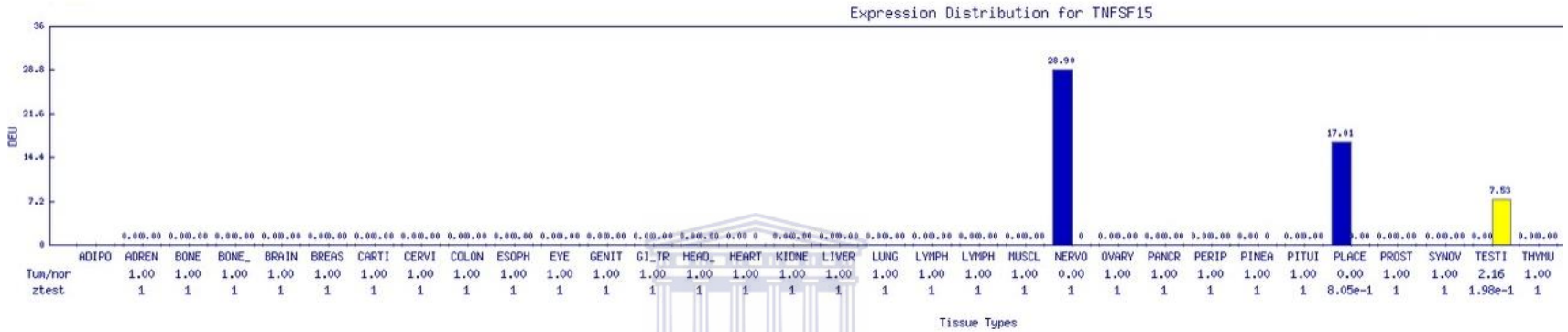


Figure 2.15: Expression profile for TNFSF15 from GeneHUB-GEPIS . Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow. There was no expression profile available for the gene in the TiGER database.

B

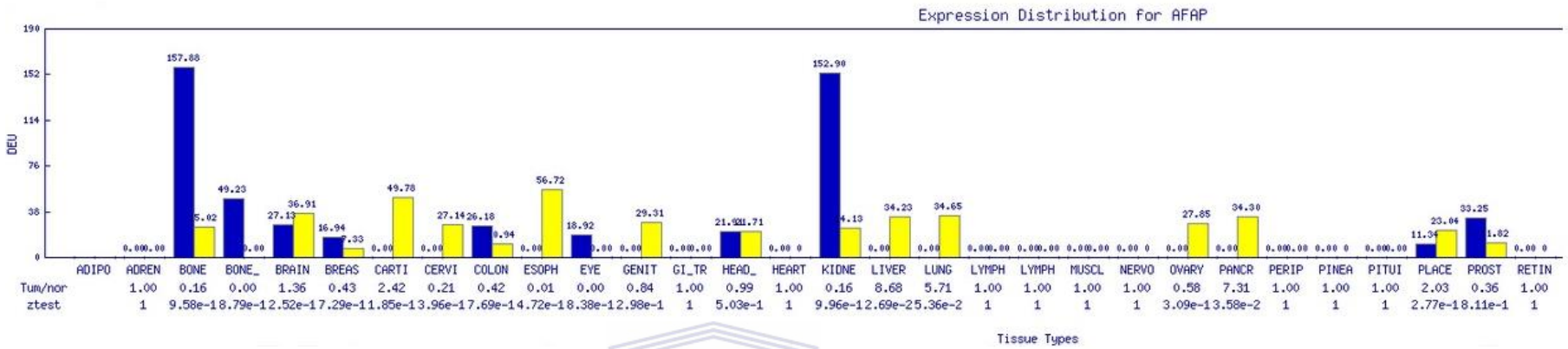


Figure 2.16: Expression profile for AFAP1 from GeneHUB-GEPIS. Normal tissue expression is shown in blue; over expression in tumour tissue is shown in yellow. There was no expression profile available for the gene in the TiGER database.

2. Gene list

Table 3: Tabular representation of the 502 target genes identified using *in silico* methods.

SGIP1	CLHC1	SGIP1	ANP32B	MSH6
E2F5	SAMD5	E2F5	MDM4	PPP4R4
ADAM22	LINC00692	TFAP2B	ANXA1	ITGA4
FOXN2	ERO1LB	TLE3	BRINP3	POGK
TOX	TMSB4Y	PACS1	UBE2E3	PLSCR1
KCMF1	OR6B2	MECP2	CYP4F31P	C3orf58
GABRB1	TWF1	FOXC1	TNKS	MAPK6
MED13	TMSB4X	KIAA1644	RP11-664D7.4	ZFP36L1
RAD23B	CRYM	AFAP1	GEMIN2	PPARG
YWHAZ	ATP5SL	EN1	TMEM74B	PTPN4
PCDHB16	TMED5	TMEM63B	TMEM209	FAM46A
PRKACB	RP11-379H8.1	ZNRF2	BCHE	TSPAN13
SLC4A10	KRTAP20-2	PIK3R2	CTPS2	RAPGEF6
WWTR1	AF131215.5	DPYSL5	COX7C	VAPA
SFMBT2	ADAM2	KMT2D	SNAP25	TMX3
NIPSNAP3A	C19orf59	HCFC1	KERA	VEZF1
C18orf25	OR2C3	MYRF	MMP21	HIC2
CSNK1G3	CTD-3148I10.1	SLC6A17	PSMD4	CLOCK
ENOX2	FAM60A	BSN	DCDC2B	SH3KBP1
KLRC4	SLC39A6	ITGA3	RBM22	CLDN18
CHD1	ATP6AP2	MBD6	OR13J1	CSNK1G3
CHCHD3	RSPH4A	CES4A	C7orf71	NEDD1
ZNF236	SRFBP1	SHANK2	WDR83	GOLPH3
HNRNPA0	PLK2	CADM4	VIMP	ADORA2B

XPO1	RGS4	SZRD1	NDUFC1	SLAIN2
RANBP9	B3GNT2	TMEM222	ZNF572	TMED7
MEX3C	GPATCH2L	PRKACA	MCOLN3	STXBP4
FAM120A	ACTR6	CNIH2	HTR7	ZBTB44
NWD2	AL353698.1	TMEM151A	HSD3B1	NUP98
GANAB	MTRNR2L8	DYSF	EIF4ENIF1	SLC31A2
EPHA5	RNF146	TCP11L1	DNTT	CAT
AKIRIN2	FAM24A	MINK1	ZNF460	UQCRH
CYTIP	GNA14	MOCS3	CLDN12	SLC31A2
KIAA2022	NOTCH2NL	STAC2	NXPH1	TPD52
MAT2B	FAM216B	CNTNAP1	PTS	MBD1
PPP2R3A	PPM1K	XYLB	TMEM241	TRMT10A
TNFSF13B	MED11	COTL1	ETNK1	AL353791.1
CNOT2	CYGB	CYP26B1	TTI1	MAPK6
POMGNT1	FYB	LRRC59	TMEM41B	ASNA1
SUB1	ADH7	PCDHB11	AC012123.1	PLK2
GDA	ZNF449	SLC9A3R2	TBCA	CTSO
ZNF148	CDIPT	PKNOX2	RAB38	MYLIP
KCNT2	BTLA	CBFA2T3	LILRB1	C11orf71
MGP	VMA21	CREB1	SSPN	SAMD9
NECAP2	SLC39A10	ELK1	ARMC2	TYSND1
ANO4	PPBP	CALR	ATF4	MLTK
MFHAS1	IMPG1	HEYL	MAEA	IMPACT
GIGYF2	MTRNR2L3	SLC7A8	RBMS1	KCNE1L
ATL1	TCEB1	KDM2A	URM1	CAT
SLIT2	KMO	SREBF2	UBE2E3	FGF10

AMMECR1L	CACNB2	TMEM194A	CD69	
CCDC67	MTUS1	STK38L	C3orf58	ZMAT2
TUBG1	LIX1L	DYNC1I1	AC011755.1	KCNJ6
TAPT1	NCOA1	SGIP1	TECRL	OTOR
ARPP19	PSME3	E2F5	ZNF667	PPP2R5E
CXADR	LARP1	NOVA1	ARPP19	NRXN2
NIPBL	IGSF11	BRINP3	B3GALNT1	ZBTB47
TSPAN6	CELF3	HMGCR	TNFSF13B	USF2
AGO3	DIRAS1	VPS13D	GSDMC	CSF2RB
ZNRF2	ELN	USP6NL	ZBED5	UBAP2L
G2E3	RSU1	RLIM	NPY1R	LRRC28
EIF5	GAS7	PANK3	RP11-766F14.2	SLC28A1
RAB18	FOXP4	HTR7	GJE1	MLLT6
LMAN1	HMP19	SDC2	KLRC4	CELSR2
ZAK	NLGN2	TEAD1	GABRB1	BCDIN3D
IPPK	FAM84B	CTCF	DDAH1	CELF5
GSDMC	SOX13	RBMS1	MMP8	PRND
ZFHX3	SMARCC2	USP32	FAM104B	ZFP36L1
HECW1	GPRC5A	SOCS7	OTC	ST3GAL2
REV3L	HNRNPU	SEMA4G	TRIM6	CCDC102B
CDK12	C6orf141	ETS1	LYRM5	WEE1
SBNO1	DYRK1A	RAB1A	POLR2D	SLC8A2
ACSL3	NFIX	PSD3	TMEM69	C20orf112
COG5	HRK	BTF3L4	JAGN1	DNAJC14
EPHA4	FZD7	SV2A	NLRP14	WDFY3
RNF139	MAT1A	TOX3	FOXO1	SLC8A1

RBM39	IGSF3	RHOT1	MUC4	DESI2
MFAP3	CLIP2	PITPNB	VAMP2	DCLK3
ADO	ZFC3H1	FAM199X	PDE4A	PEAR1
ST3GAL6	KDM5C	CLDN12	SPRED3	
ZDHHC21	LOC101927910	TIMP3	FAM131B	
SPTBN4	FCHO2	SERPINB7	C1orf198	
PTHLH	ZMYM4	FOXE1	RARA	
CALD1	YLPM1	HOXC11	PRELP	
NR1D1	KIAA1549L	UNC13A	EIF4G1	
DAAM2	SUMO2	ASIC1	TACC1	
DAB2IP	BTBD10	PXN	GLDN	
C9orf57	TMEM168	PCSK2	PBX1	
PLA2G2D	MAP9	MYL12A	PACSIN1	
SPRED2	ADAMTS6	ALX4	SPIN3	
CNST	TCF7L1	UBTF	ATXN7L3	
MTSS1L	OTUD4	LAMC3	CDK18	
ATCAY	RSBN1	TMEM184B	ZNF385B	
PVALB	EIF4ENIF1	WHSC1	CPLX2	
SPRR1B	SEMA3E	LCLAT1	DHRS11	
RANBP10	EXOC8	ZFP36	CTNND1	

Appendix B

Chapter 3 supplementary information.

Melting curves

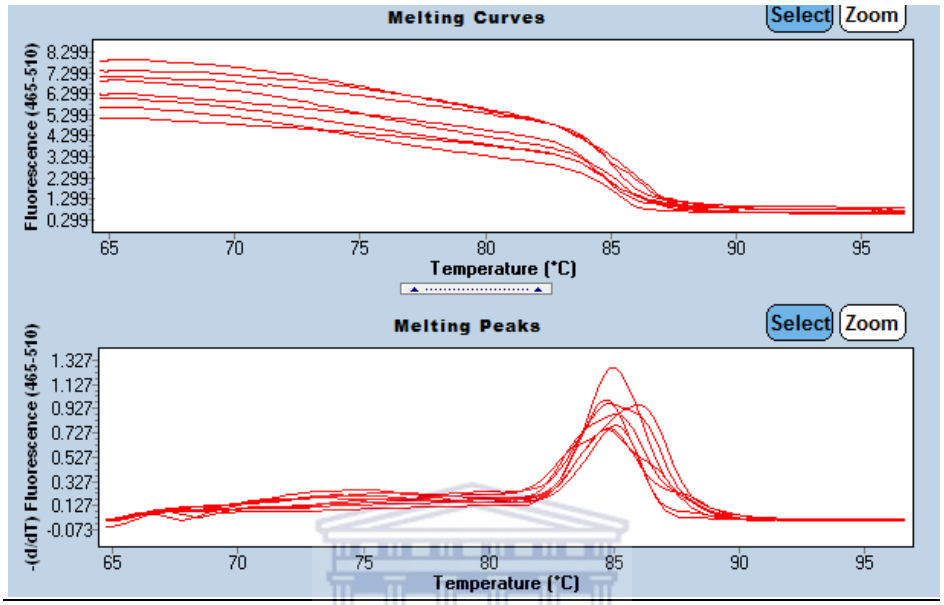


Figure 3.8: Melting curve and melting peak of miR1 in MCF 7 cells.

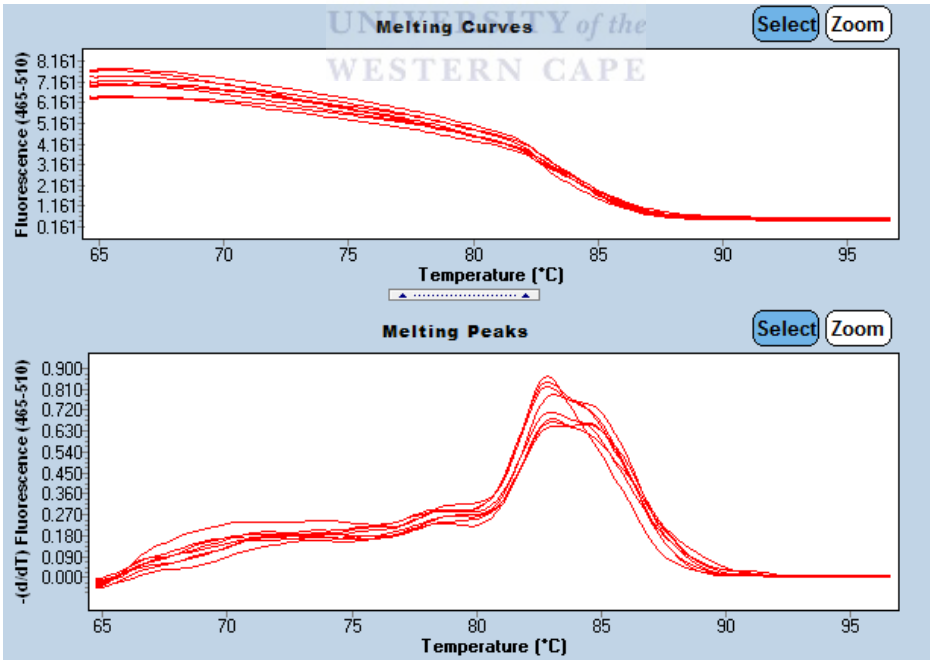


Figure 3.9: Melting curve and melting peak of miR2 in MCF 7 cells.

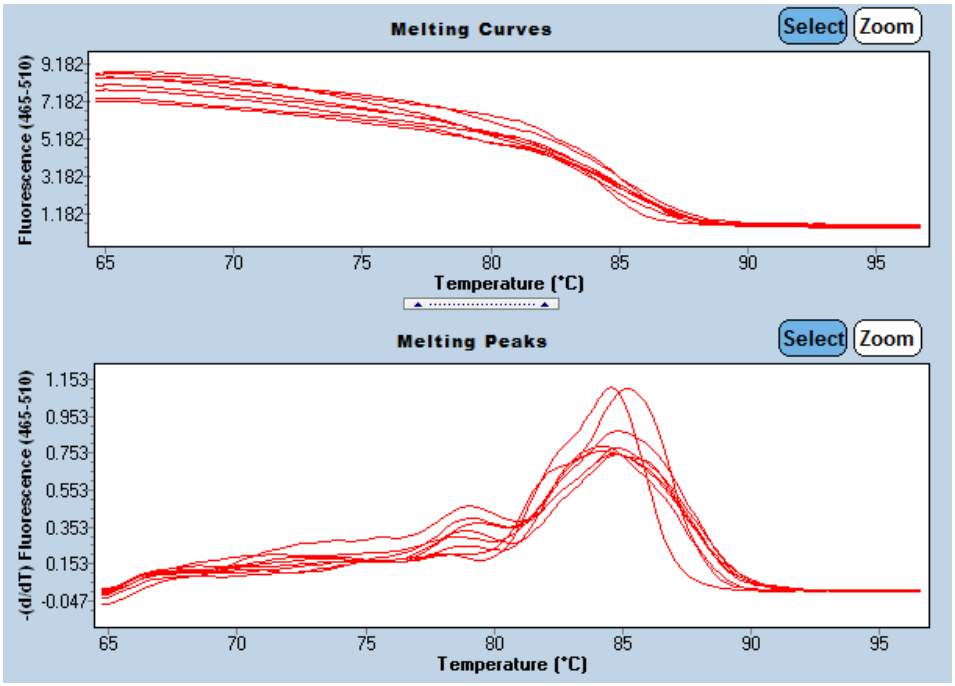


Figure 3.10: Melting curve and melting peak of miR3 in MCF 7 cells.

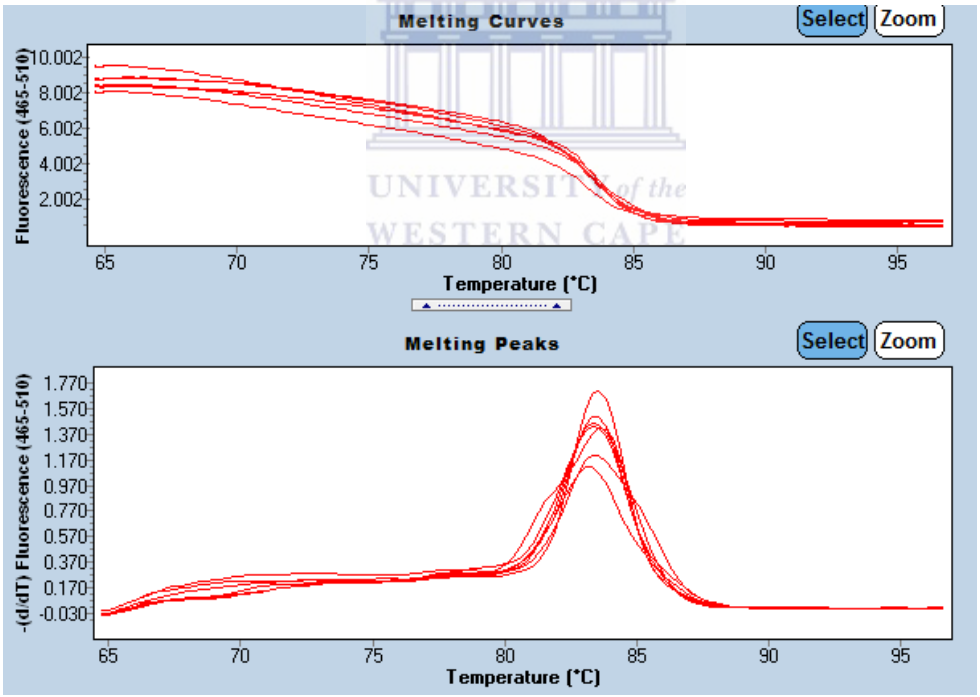


Figure 3.11: Melting curve and melting peak of miR4 in MCF 7 cells. A sharp peak is seen at the T_m of 84 ºC.

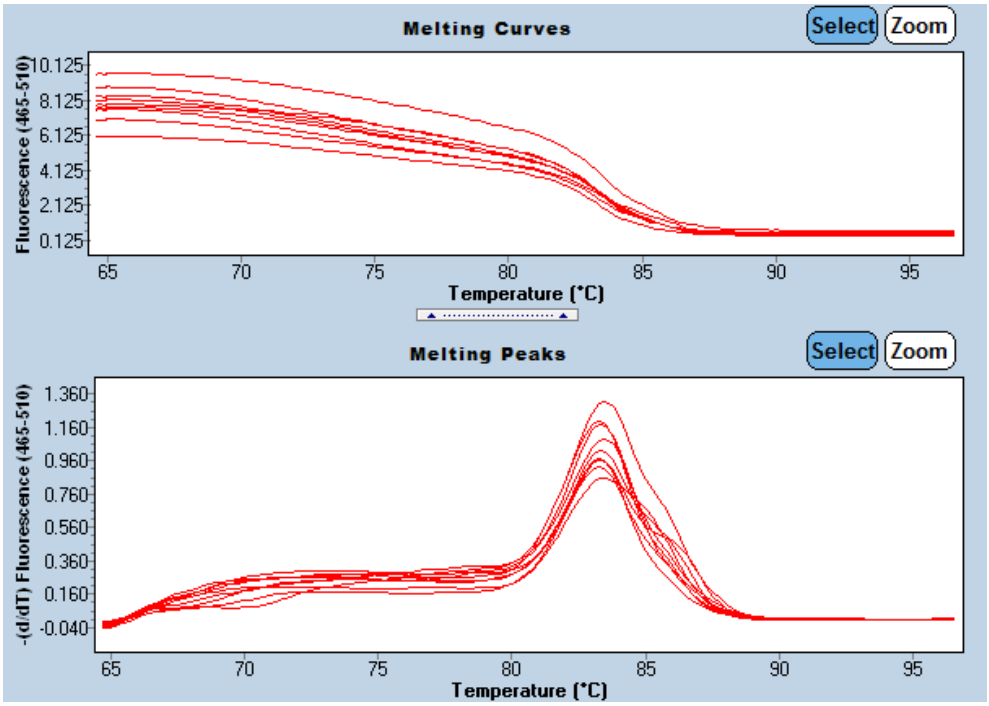


Figure 3.12: Melting curve and melting peak of miR5 in MCF 7 cells.

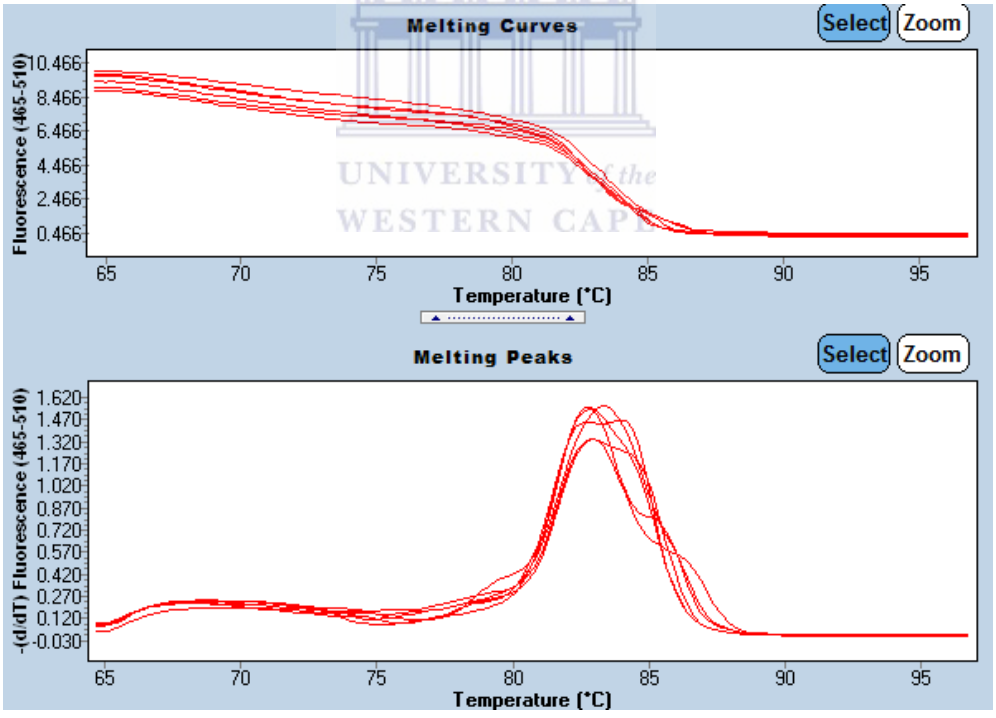


Figure 3.13: Melting curve and melting peak of miR1 in LNCaP cells.

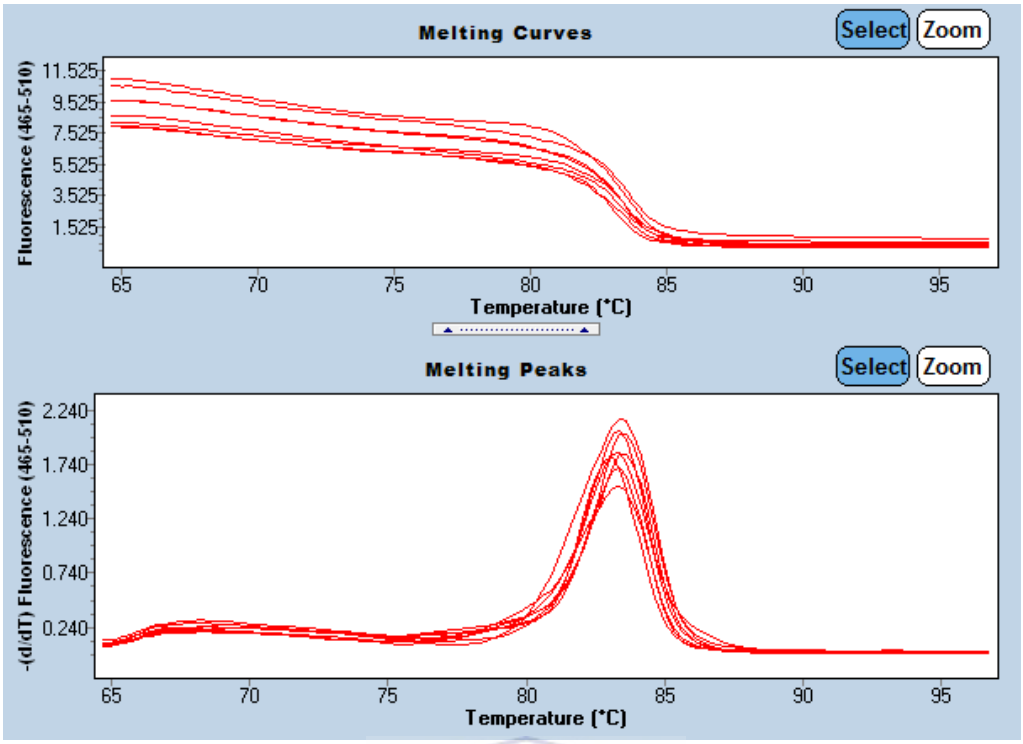


Figure 3.14: Melting curve and melting peak of miR2 in LNCaP cells.

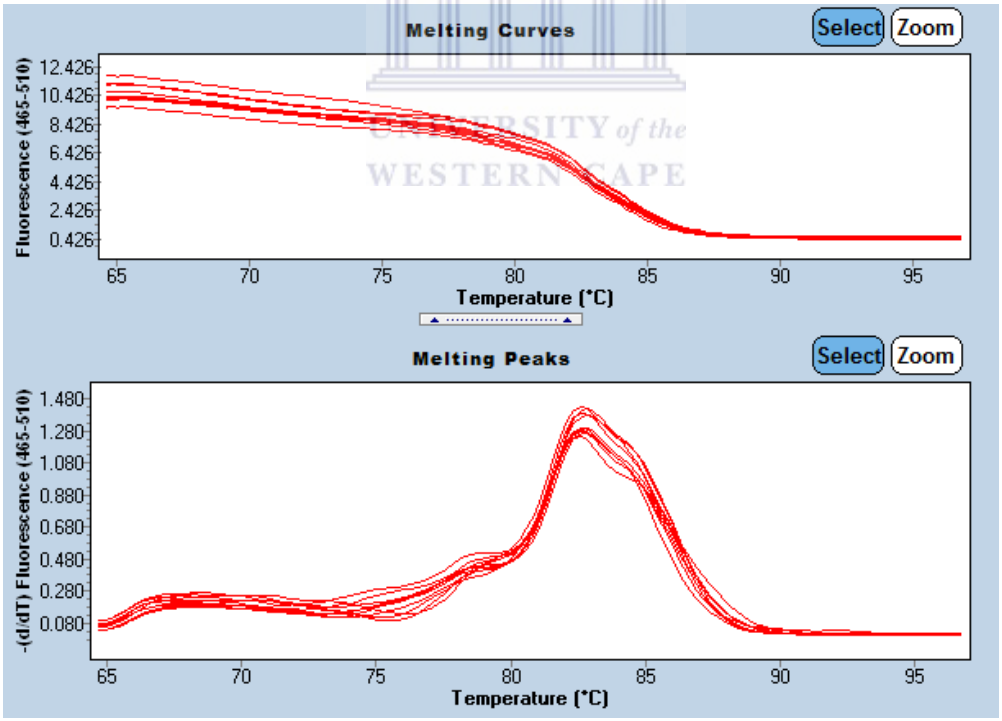


Figure 3.15: Melting curve and melting peak of miR3 in LNCaP cells.

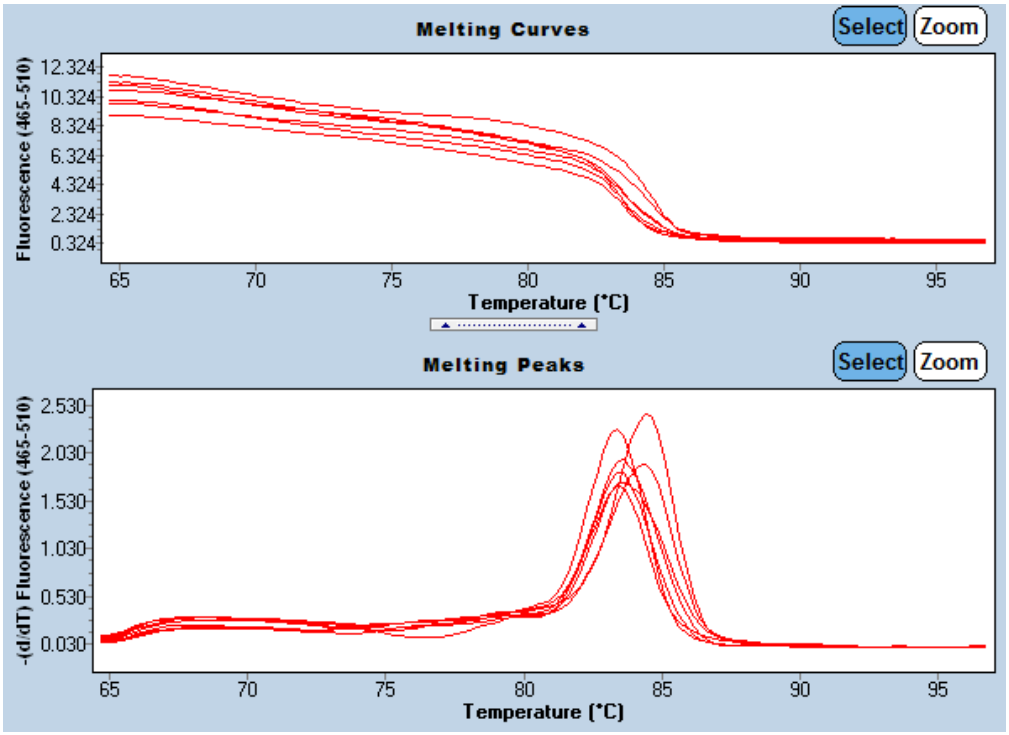


Figure 3.16: Melting curve and melting peak of miR4 in LNCaP cells.

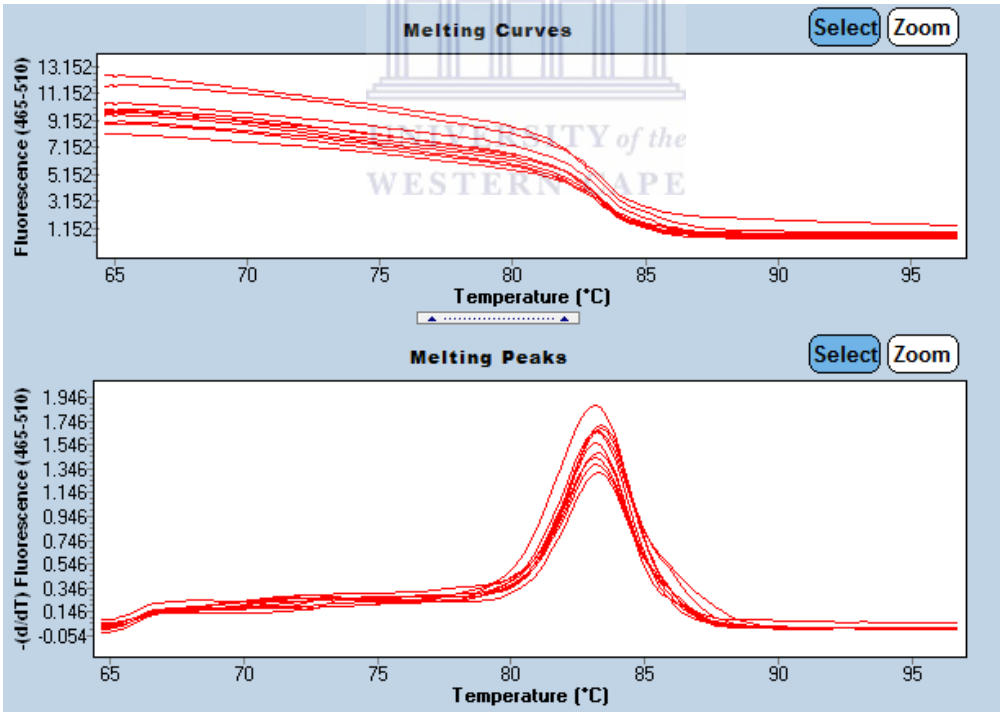


Figure 3.17: Melting curve and melting peak of miR5 in LNCaP cells.

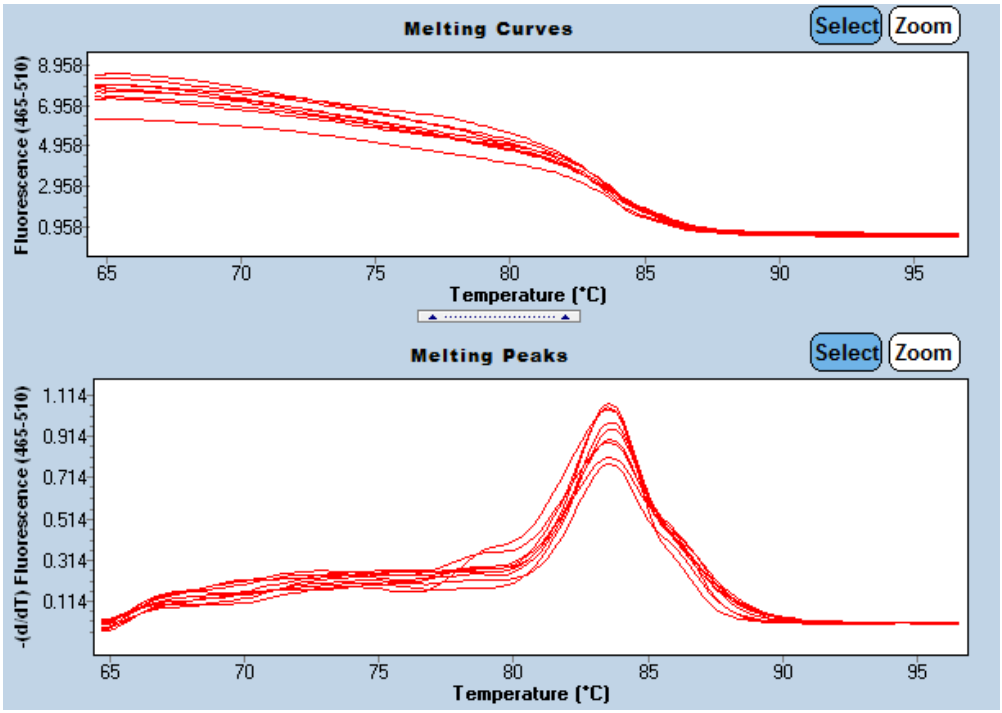


Figure 3.18: Melting curve and melting peak of miR2 in BPH1 cells.

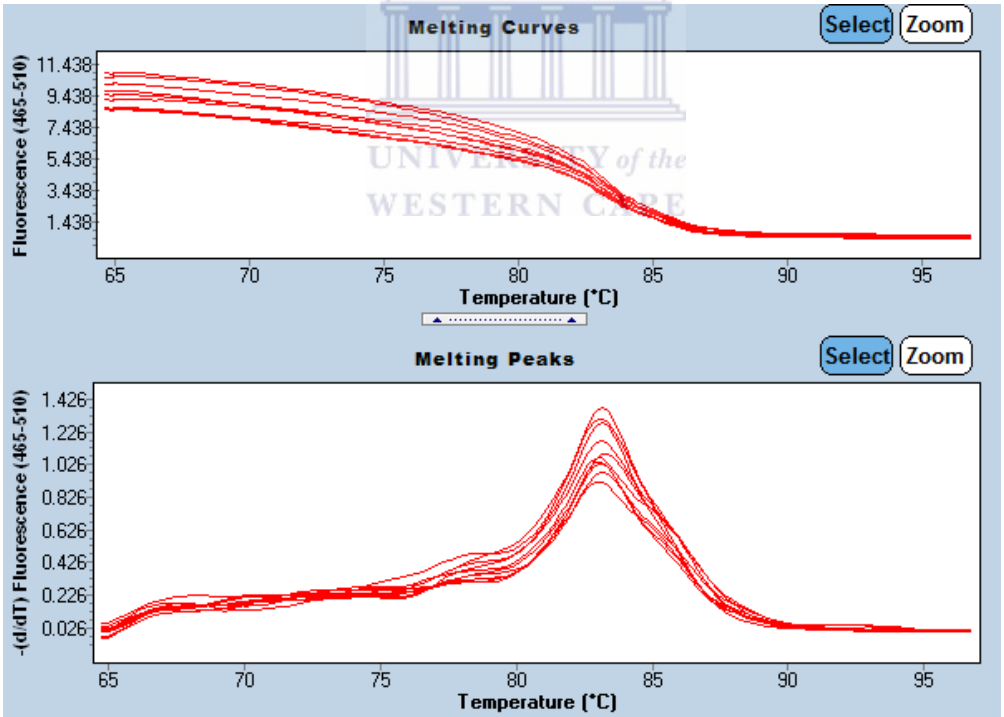


Figure 3.19: Melting curve and melting peak of miR3 in BPH1 cells.

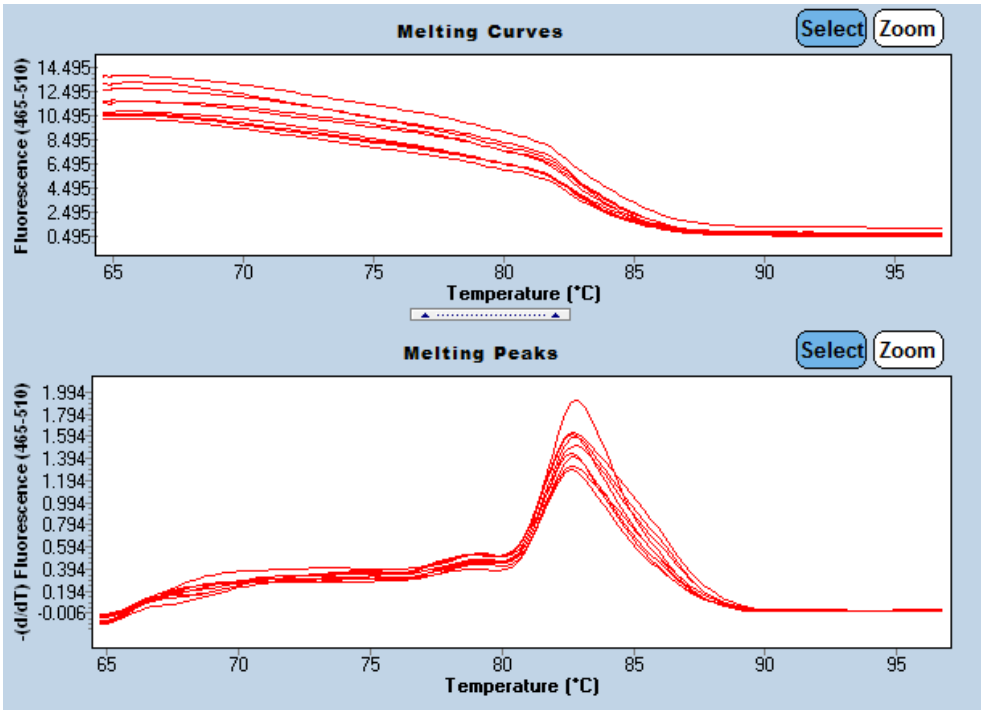


Figure 3.20: Melting curve and melting peak of miR4 in BPH1 cells.

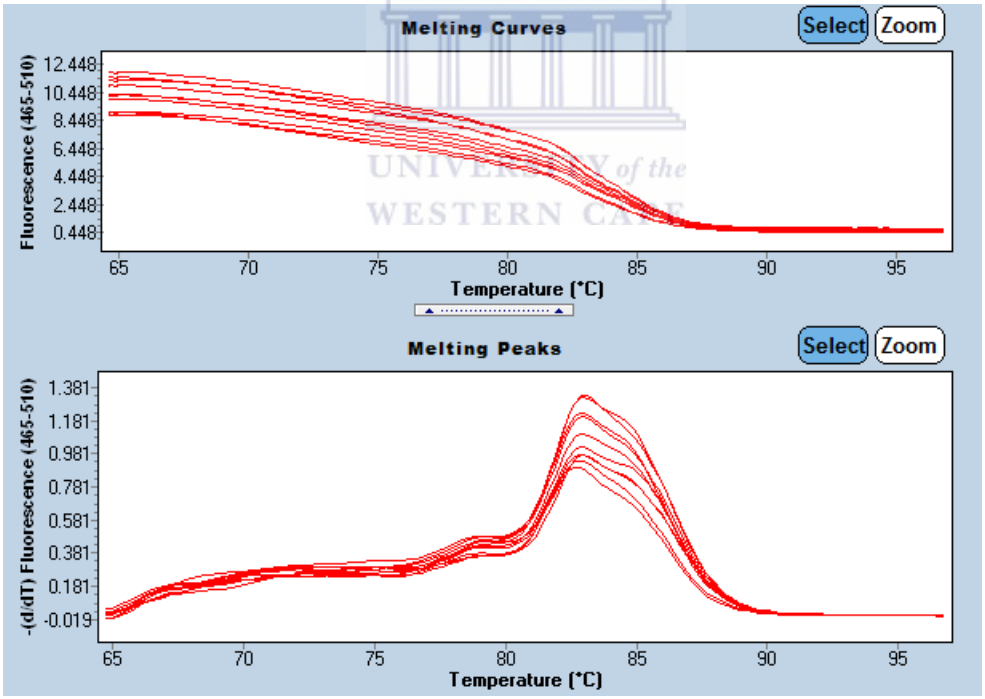


Figure 3.21: Melting curve and melting peak of miR5 in BPH1 cells.

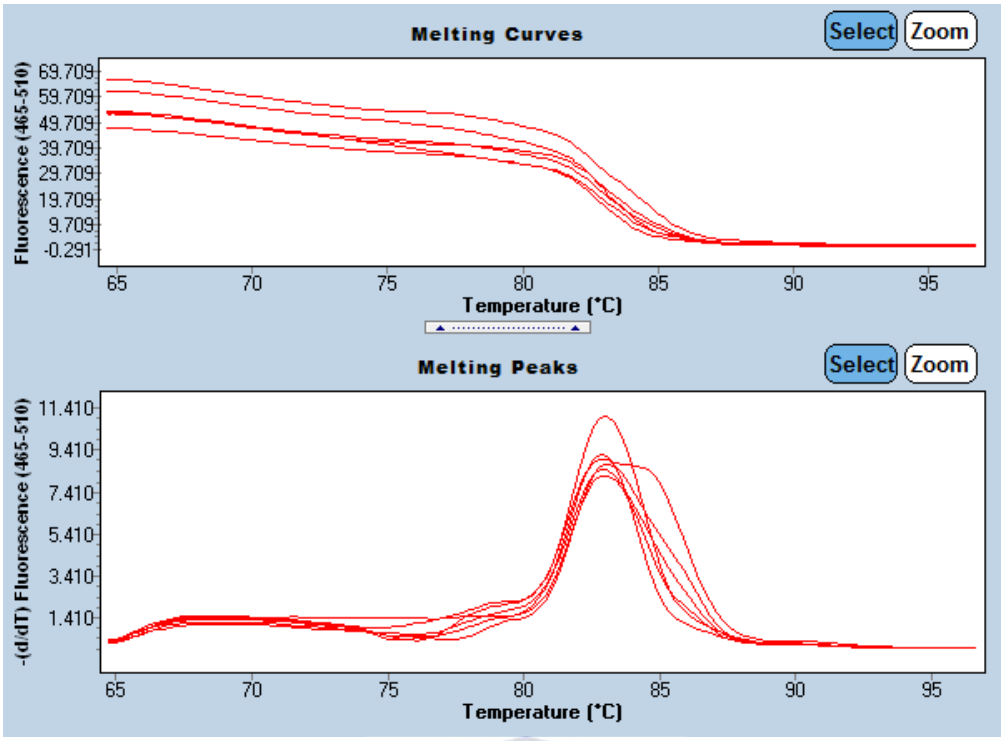


Figure 3.22: Melting curve and melting peak of miR1 in PNT1a cells

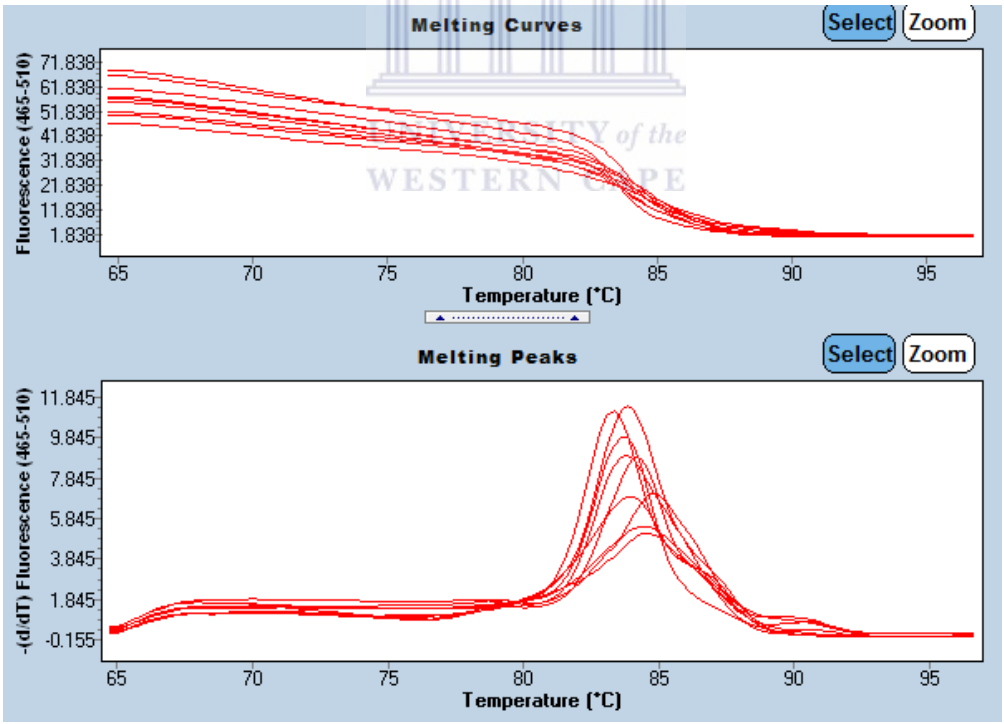


Figure 3.23: Melting curve and melting peak of miR2 in PNT1a cells.

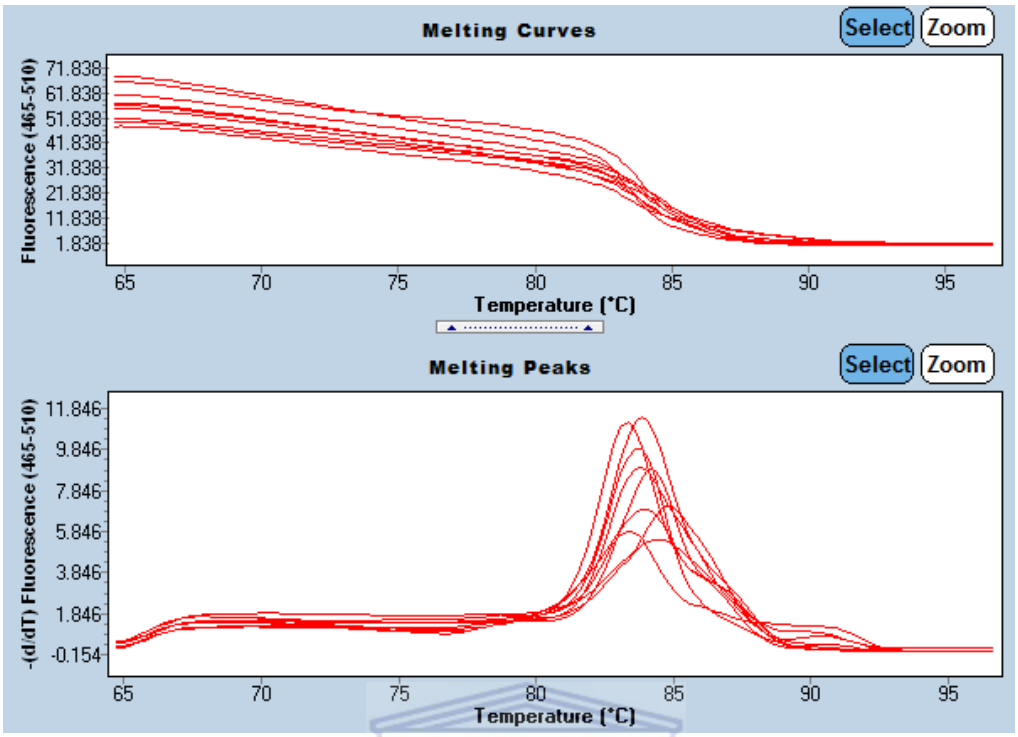


Figure 3.24: Melting curve and melting peak of miR3 in PNT1a cells.

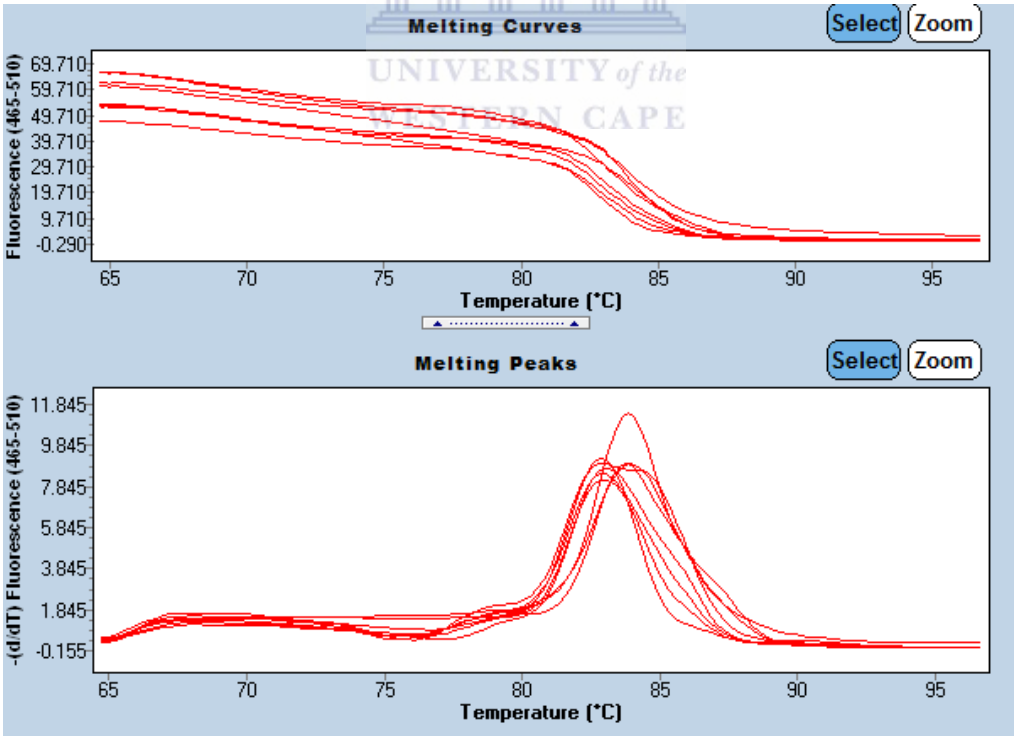


Figure 3.25: Melting curve and melting peak of miR4 in PNT1a cells.

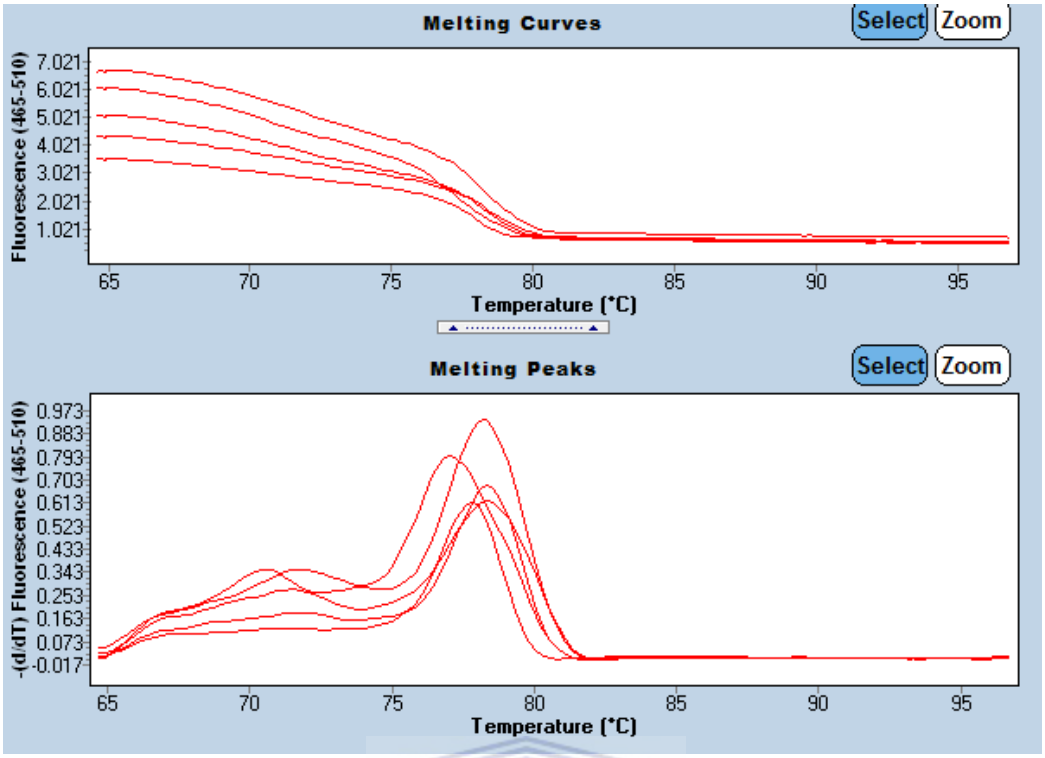


Figure 3.26: Melting curve and melting peak of miR5 in PNT1a cells.

References

1. Abdullah-Sayani, A., Bueno-de-Mesquita, J.M., van de Vijver, M.J. (2006). Technology Insight: tuning into the genetic orchestra using microarrays—limitations of DNA microarrays in clinical practice. *Nature Clinical Practice Oncology*, 3(9):501-516.
2. Adrain, C., Creagh, C.M., Martin, S.J. (2003). Apoptosis-associated release of Smac/DIABLO from mitochondria requires active caspases and is blocked by Bcl-2. *EMBO J*. 2001. 20(23): 6627–6636.
3. Agarwal, V., Bell. G.W., Nam, J. W., Bartel, D., P. (2015). "Predicting effective microRNA target sites in mammalian mRNAs.
4. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7th edition. New York, NY: *Springer Science Business Media*, 2010.
5. Ang, C.W., Dawson, R., Hall, C., Farmer, M., (2008). The diagnostic value of digital rectal examination in primary care for palpable rectal tumour. *The Journal of Colorectal Disease*, 10 (8):789-92.
6. Ardekani, A.M and Naeini, M.M. (2010). The Role of MicroRNAs in Human Diseases. *Advanced Journal of Medical Biotechnology* 2(4):161-179.
7. Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F. (2011). Argonaute 2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences U S A.*
8. Atkinson, A.J., Colburn, W.A., DeGruttola, V.G., DeMets DL. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Biomarkers Definitions Working Group. *Clinical Pharmacology & Therapeutics* 69:89–95.

9. Babb, C., Urban, C., Kielkowski, D. and Kellett. P. (2014). Prostate Cancer in South Africa: Pathology Based National Cancer Registry Data (1986–2006) and Mortality Rates (1997–2009). *Prostate Cancer* 39:12-27.

10. Bartel, D. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.

11. Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136: 215–33.

12. Bartel, D. P. Chen, C. Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature. Reviews. Genetics.*5:396-400.

13. Basch, E., Oliver, T.K., Vickers, A. (2012). Screening for prostate cancer with prostate-specific antigen testing: American Society of Clinical Oncology provisional clinical opinion. *Journal of Clinical Oncology*. 30:3020-3025.

14. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. (2012). GenBank. Nucleic Acids Research, 40:48-53.

15. Berman H, Henrick K, Nakamura H. (2003). Announcing the worldwide Protein Data Bank. Nature Structural and Molecular Biology, 10:990-1203. *Bioinformation*, 6(3): 95–99.

16. Boffetta, P., (2010) Biomarkers in cancer epidemiology: an integrative approach. *Carcinogenesis.* 31:121-6.

17. Bohnsack, M.T., Czaplinski, K., Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA.10:185-91.

18. Bostwick, D.G., and Brawer, M.K. (1987). Prostatic Intraepithelial neoplasia and Early Invasion in Prostate Cancer. *Cancer* 59:788–794.

19. Brase, J.C, Johannes, M., Schlomm, T., Falth, M., Haese, A., Steuber, T. (2011). Circulating miRNAs are correlated with tumour progression in prostate cancer. *International Journal of Cancer*, 128:608-616.

20. Bushati, N., and Cohen, S.M. (2007). microRNA Functions. *Annual Review of Cell and Developmental Biology*, 23:175-205.

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

22. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukaemia. *Proceedings of the National Academy of Sciences USA*, 99: 15524-15529.

23. Cancer Biomarkers: Minimal and Non-invasive Early Diagnosis and Prognosis - Google Books. 2015. Cancer Biomarkers: Minimal and Non-invasive Early Diagnosis and Prognosis - Google Books. [ONLINE] Available at: https://books.google.co.za/books?id=8IBcAgAAQBAJ&pg=PA428&lpg=PA428&dq=Chen+et+al.+Oncologist.+2012+biomarkers&source=bl&ots=vH7qfKZCqP&sig=NCyB1e5fXJTzCXK-UvsLhClChB4&hl=en&sa=X&ved=0CDUQ6AEwA2oVChMI17vt3sf5yAIVAlIaCh0HEw5X#v=onepage&q=Chen%20et%20al.%20Oncologist.%202012%20biomarkers&f=false [Accessed 01 November 2015].

24. Cannell, I.G., Kong, Y.W., Bushell, M. (2008) How do microRNAs regulate gene expression? *Biochemical SocietyTransactions* 36:1224–1231.

25. Carlsson, J., Davidsson, S., Helenius, G., Karlsson, M., Lubovac, Z., Andrén, O., Olsson, B., Klinga-Levan, K. (2011). A miRNA expression signature that separates between normal and malignant prostate tissues. *Cancer Cell International*, 11**:**14-19.
26. Carter, H.B., Albertsen. P.C., Barry, M.J., Early detection of prostate cancer: AUA guideline. www.auanet.org/common/pdf/education/clinical-guidance/Prostate-Cancer-Detection.pdf.
27. Chai, J., Chunying, Du., Wu, J.W., Kyin, S., Wang, X., Shi, Y., (2000). Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature* 406, 855-862.
28. Chan, J.A., Krichevsky, A.M., Kosik, K.S. (2005). MicroRNA-21 is an Antiapoptotic Factor in Human Glioblastoma cells. *Cancer Research 65:6029*-6033.
29. Chang, B. L., Spangler. E., S. Gallagher (2011) Validation of genome-wide prostate cancer associations in men of African descent, *Cancer Epidemiology, Biomarkers & Prevention*, 20:23–32.
30. Cheng, H.L., Huang H.J., Ou B.Y., Chow N.H., Chen Y, W., Tzai, T.S., Wu, C.J., Chen, C.H., (2011). Urinary CD14 as a potential biomarker for benign prostatic hyperplasia – discovery by combining MALDI-TOF-based biostatistics and ESI-MS/MS-based stable-isotope labelling. *Proteomics Clinical Applications* 5:121-132.
31. Chu, L.W., Ritchey. J., Devesa. S., Quraishi, S.M., Zhang, H. and Hsing, A.W., (2011). Prostate cancer incidence rates in Africa. *Prostate Cancer*, 94:70-78.

32. Colotta, F., Allavena, P., Sica, A., Garlanda, C., Mantovani, A. (2009). Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*, 30:1073–1081.

33. Corcoran, D.L., (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoSOne*. 4:52-69.

34. Croce C.M., Schetter, A.J., Leung, S.Y., Sohn, J.J., Zanetti, K., Bowman, E.D., *et al*., (2008). MicroRNA Expression Profiles Associated with Prognosis and Therapeutic Outcome in Colon Adenocarcinoma. *American Medical Association*, 299(4):425-436.

35. Croce, C.M., Garzon, R., and Calin, G.A. (2009). MicroRNAs in Cancer. *Annual Review of Medicine*, 60: 167-179.

36. Curado, M. P., Edwards, B., Shin, H.R., (2007). Cancer Incidence in Five Continents. *IARC Scientific*, 9:15-20.

37. Davis, B.N., Hilyard, A.C., Lagna, and G., Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*, 3;454(7200):56-61.

38. De Bock. M., de Seny. D., Meuwis. M.A, Chapelle. J, Louis. E, Malaise. M, Merville. M.P, and Fillet. M, 2010, Challenges for Biomarker Discovery in Body Fluids Using SELDI-TOF-MS *Journal of Biomedicine and Biotechnology*, 1-15.

39. Denli, A. M.; Tops, B. B.; Plasterk, R. H.; Ketting, R. F.; Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 2004, *432*, 231-235. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Research*, 66(12):6097–104.

40. Dennis, G. Jr., Sherman., B.T., HosackD. A., Yang. J., Gao, W., Lane, C.H., Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Journal of Genome Biology*, Volume 4:9.
41. Deo, A., Carlsson, J., Lindlöf, A. (2011). How to choose a normalization strategy for miRNA quantitative real-time (qRT-PCR) arrays. *Journal of Bioinformatics and Computational Biology*, 9:79-85.
42. Derveaux, S., Vandesompele, J., Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods*, 50:227-30.
43. Descotes. J-L., Legeais D, Gauchez A-S, Long J-A, Rambeaud J-J. PSA (2007). Measurement following prostatectomy: an unexpected error. *Anticancer Research* ,27:1149-1150.
44. Draghici, S., Khatri, P., Eklund, A.C., Szallasi, Z., (2008). Reliability and reproducibility issues in DNA microarray measurements. *Trends in Genetics*, 24 (12) 583-90.
45. Drucker, E., Krapfenbauer (2013). Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *The EPMA Journal*, 4:7-12.
46. Du, T., Zamore, P. D. (2005). microPrimer: the biogenesis and function of microRNA. *Development*, *132*, 4645-4652.
47. Dyche, D.J., Ness, J., West, M., Allareddy, V., Konety, B.R., (2006). Prevalence of prostate specific antigen testing for prostate cancer in elderly men. *Journal of Urology*, 175(6):2078-82.
48. Endo, T., Uzawa, K., Suzuki, H., Tanzawa, H., Ichikawa, T. (2009). Characteristic gene expression profiles of benign prostatic hypertrophy and prostate cancer. *International Journal of Oncology* 35:499-509.

Epidemiology and Prevention. *IARC Scientific*, 153-60.

49. Epstein, J.I. (1998). "Pathologic Features that Predict Progression of Disease Following Radical Prostatectomy," in eds. Foster, C.S., and D.G. Bostwick, Pathology of the Prostate. Philadelphia, Pa.: W.B. Saunders Company, 228-244.
50. Evans. S., C. Metcalfe, B. Patel *et al*. (2010). Clinical presentation and initial management of Black men and White men with prostate cancer in the United Kingdom: the PROCESS cohort study. *British Journal of Cancer*, 102 (2):249–254.
51. Fabbri M. (2010). miRNAs as molecular biomarkers of cancer. *Expert Review of Molecular Diagnostics,* (4):435-44.
52. Federman, D.G., Pitkin, P., Carbone, V., Concato, J., Kravetz, J.D. (2014). Screening for prostate cancer: are digital rectal examinations being performed? *The National Institutes of Health,* 42(2):103-7.
53. Ferlay, J., H. Shin., F. Bray., D. Forman., C. Mathers., and D. Parkin, GLOBOCAN, 2008, Cancer Incidence and MortalityWorldwide, IARC Cancer Base No. 10, *International Agency for Research on Cancer*, Lyon, France, 2010.
54. Fraga, D., Meulia, T., and Fenster, S. 2008. Real-Time PCR. *Current Protocols Essential Laboratory Techniques,* 10.3.1-10.3.34.
55. Framework for Biomarker and Surrogate Endpoint Use in Drug Development. Woodcock J; http://www.fda.gov/ohrms/dockets/ac/04/slides/2004-4079S2_03_Woodcock.ppt.
56. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acid Research,* Volume 41.

57. Francis, D.B., Kozlov, M., Chavez, J., Chu, J., Malu, S., Hanna. M., Cortes, P. (2014). DNA Ligase IV regulates XRCC4 nuclear localization. J*ournal of DNA Repair*, 21:36-42.

58. Friedman, R.C., Farh, K.K., Burge, C.B., Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of miRNAs. *Genomic Research*, 19: 92–105.

59. Funari, V.A., Voevodski, K., Leyfer, D., Yerkes, L., Cramer, D., Tolan, D.R. (2010) Quantitative Gene Expression Profiles in Real Time from Expressed Sequence Tag Databases. *Gene Expression Patterns Journal*, 14(6):321-36.

60. Garzon, R., Marcucci, G., Croce, C.M. (2014). Targeting MicroRNAs in Cancer: Rationale, Strategies and Challenges. Drug Discovery 9, 775-789.

61. Geneva, W.H.O. Global health risks: Cancer fact sheets. (2015).

62. Gharaee-Kermani M and Macoska, J.A. (2013). Promising molecular targets and biomarkers for male BPH and LUTS. *Current Urology Reports*, 14(6):628-37.

63. Gleason, D., (1966) "Classification of prostatic carcinomas," Cancer Chemotherapy Reports 50 (3): 125–128.

64. Gong, H., Liu., C.M., Liu, D.P., Liang, C.C., (2005). The role of small RNAs in human diseases: potential troublemaker and therapeutic tools. *Medical Research Reports*, 25(3):361-81.

65. Good. D, Thongboonkerd. V., Novak. J., Bascands. J.P., Schanstra. J.P, Coon. J.J., Dominiczak. A., and XMischak. H. (2007). Body Fluid Proteomics for Biomarker Discovery: Lessons from the Past Hold the Key to Success in the Future. *Journal of Proteome Research*, 6 (12): 4549-4555.

66. Grant, W., Cannon, B.S., Getzenberg R, H. (2012). Biomarkers for Benign Prostatic Hyperplasia Progression. *Current Urology Reports*. 9(4): 279–283.

67. Gregory, R. I., Chendrimada, T. P., Cooch, N.; Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and post-transcriptional gene silencing. *Cell*, *123*, 631-640.

68. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36, D154-158.

69. Gupta, R.T., Kauffman, C.R., Polascik, T.J., Taneja, S.S., Rosenkrantz, A.B. (2013). The state of prostate MRI in 2013. *Oncology*, 27:262-270.

70. Haj-Ahmad, T.A., Abdalla, M.A., Haj-Ahmad, Y. (2014). Potential Urinary miRNA Biomarker Candidates for the Accurate Detection of Prostate Cancer among Benign Prostatic Hyperplasia Patients. *Journal of Cancer* 29:5(3):182-9.

71. Hanahan. D and Weinberg. R.A. (2000). Hallmarks of Cancer: The Next Generation. *Cell*, 144:5:646–674.

72. Hanash, S. (2011). Progress in Mining the Human Proteome for Disease Applications. *OMICS: a Journal of Integrative Biology*, 15: 133–139.

73. Hancock, T., Takigawa, I., Mamitsuka, H. (2013). Identifying pathways of coordinated gene expression. Methods in Molecular Biology, 939:69-85.

74. Hanke, M., Hoefig, K., Merz, H., Feller, A.C., Kausch, I., Jocham, D., Warnecke, J.M., Sczakiel, G. (2010). A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. *Urology Oncology*, 28(6):655-61.

75. He, L., Thomson, J.M., Hernando-Monge, E., Mu, D., Goodson, S., Hannon, G.J., Hammond, S.M., Lowe, S.W., Cordon-CArdo, C. (2005). A microRNA polycistron as a potential human oncogene. *Nature* 435(7043):823–8.

76. Heid, C. A., Stevens, J., Livak, K.J., Williams, P.M. (1996). Real-time quantitative PCR. *Genome Research*, 6:986-994.

77. Heidenreich, A., Bellmunt, J., Bolla, M., Joniau, S., Mason, M., Matveev, V., Mottet, N., Schmid, H.P., van der Kwast, T., Wiegel, T., Zattoni, F. (2011). EAU guidelines on prostate cancer. Part I: screening, diagnosis, and treatment of clinically localised disease. *European Association of Urology*, 59:61-71.
78. Heyns. C. F., M. Fisher, A. Lecuona, and A. van der Merwe. (2011). Prostate cancer among different racial groups in the Western Cape: presenting features and management," *South African Medical Journal*, 101(4):267–270.
79. HMGA2 Maintains Oncogenic RAS-Induced Epithelial-Mesenchymal Transition in Human Pancreatic Cancer Cells. *The American Journal of Pathology*, 174 (3).
80. Horwich A, Parker C, Bangma C, Kataja V. (2010). On behalf of the ESMO Guidelines Working Group. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annual Journal of Oncology*, 21(5):129-33.
81. Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research*, 42:78–85.
82. Huan, C and Freter, C. (2014). Lipid Metabolism, Apoptosis and Cancer Therapy. *International Journal of Molecular Science*, 16:1: 924-949.
83. Huang, D.W., Sherman, B.T., Lempicki, R.A, (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1-13.
84. Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, G.W., Roayaei, J., Stephens, R., Baseler, M.W., Lane, C.H., Lempicki, R.A. (2007). The DAVID

Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Journal of Genome Biology*, 8:9.

85. Huang, P.W., Lee, C.H., Lin, P.L., (2013). Classifying pathological prostate images by fractal analysis. In: Chatterjee A, Siarry P, eds. Computational Intelligence in Image Processing. Heidelberg, Germany: *Springer-Verlag Berlin Heidelberg*, 253-263.
86. Inoue, J., Ishida, T., Tsukamoto, N., Kobayashi, N., Naito, A., Azuma, S., Yamamoto, T. (2000). "Tumour necrosis factor receptor-associated factor (TRAF) family: adapter proteins that mediate cytokine signalling". *Experimental Cell Research*, 254 (1): 14–24.
87. Iorio, M.V., and Croce, C.M. (2012). microRNA involvement in human cancer. *Carcinogenesis*, 33(6):1126-33.
88. Ishiguro, H., Akimoto, K., Nagashima, Y., Kojima, Y., Sasaki, T., Ishiguro-Imagawa, Y., Nakaigawa, N., Ohno, S., Kubota, Y., Uemura, H. (2009). aPKClambda/iota promotes growth of prostate cancer cells in an autocrine manner through transcriptional activation of interleukin-6. *Proceeding of the National Academy of Science USA*, 22; 106 (38):16369-74.
89. Ishiguro, H., Akimoto, K., Nagashima, Y., Kojima, Y., Sasaki, T., Ishiguro-Imagawa, Y., Nakaigawa, N., Ohno, S., Kubota, Y., Uemura, H. (2009). aPKClambda/iota promotes growth of prostate cancer cells in an autocrine manner through transcriptional activation of interleukin-6. *Proceeding of the National Academy of Science USA*, 22; 106 (38):16369-74.
90. Istrail, S., Davidson, E. H. (2005). Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Science*, 102(14):4954-4959.

91. Jensen, T.W., Ray, T., Wang, J., Li X., Naritoku, W.Y., Han, B., Bellafiore, F., Bagaria, S.P., Qu, A., Cui, X., Taylor, C. R., Ray, P.S. (2015). Diagnosis of Basal-Like Breast Cancer Using a FOXC1-Based Assay. *Journal of the National Cancer Institute* 107:8.

92. Ji, H., Wang, J., Liu, J., Guo, J., Wang, Z., Zhang, X., Guo, L. and Yang, H. (2013). Selection of Reliable Reference Genes for Real-time qRT-PCR Analysis of Zi Geese (*Anser anser domestica*) Gene Expression. *Asian-Australasian Journal of Animal Sciences*, 26(3): 423 – 432

93. Jiang Q *et al*., (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*, 37:98–104.

94. Jiao, X., Sherman, B.T., Stephens, R., Baseler, M. W., Lane, H.C., Lempicki, R. A. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 28: 1805-1806.

95. Karazanashvili, G., Abrahamsson, P.A. (2003). Prostate specific antigen and human glandular kallikrein 2 in early detection of prostate cancer. *Journal of Urology*, 169 (2):445–457.

96. Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K *et al*., (2005). Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Science*, 96: 111-115.

97. Kim, Y. K., Kim, V. N. (2007). Processing of intronic microRNAs. *EMBO J, 26*: 775-783.

98. Kobel, M., Gilks, C.B., Huntsman, D.G. (2009). Dicer and Drosha in ovarian cancer. *New England Journal of Medicine*, 12: (11):1150-1.

99. Krallinger, M., Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome Biology*, 6(7).

100. Kulasingam, V., Diamandis E.P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practical Oncology*, (10):588-99.

101. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., (2001). Identification of Novel Genes Coding for Small Expressed RNAs *Science*, 294(554):3853-858.

102. Landgraf, P. *et al*., (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129 (7):1401–14.

103. Lange, E. M., Sarma, A. Ray *et al*., (2008). "The androgen receptor CAG and GGN repeat polymorphisms and prostate cancer susceptibility in African-American men: results from the Flint Men's Health Study." *Journal of Human Genetics*, 53(3):220–226.

104. Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294:858–862.

105. Lawrie, C.H., Gal, S., Dunlop, H.M. (2008). Detection of elevated levels of tumour-associated miRNAs in serum of patients with diffuse large B-cell lymphoma. British Journal of Haematology. 141:672–675.

106. Lee, R.C., Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294: 862–864.

107. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 3;75(5):843-54.

108. Lee, Y., Jeon, K., Lee, J. T., Kim, S., Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J., 21*, 4663-4670.

109. Li, A., and Mao, L. (2007). Evolution of plant microRNA gene families. *Cell Research*, 17: 212–218.

110. Lin, S. L., Chang, D., Wu, D. Y., Ying, S. Y. (2003). A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution. *Biochemical and Biophysical Research Communications, 310*:754-760.

111. Linsley, P.S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M.M., Bartz, S.R., Johnson, J.M., Cummins, J.M., Raymond, C.K., Dai, H., Chau, N., Cleary, M., Jackson, A.L., Carleton, M., Lim, L. (2011). Transcripts regulated by the microRNA-16 family cooperatively regulate cell cycle progression. *Molecular and Cellular Biology*, 27:2240-2252.

112. Liu, X., Yu, X., Zack D.J., Zhu, H., Qian J. (2008). TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics,* 9:271.

113. Livak, K. J. 1997. ABI Prism 7700 Sequence Detection System, User Bulletin 2. PE Applied Biosystems, Foster City, CA.

114. Lowe, S.W., Lin, A. (2000) Apoptosis in cancer. *Journal of Carcinogenesis*, 21: 485-495.

115. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck D *et al* (2005). MicroRNA expression profiles classify human cancers. *Nature*, 435: 834-838.

116. Luo, J., Solimini, N.L., Elledge, S.J. (2009). Principles of cancer therapy: Oncogene and non-oncogene addiction. *Cell*, 136: 823–837.

117. Luo, J., Solimini, N.L., Elledge, S.J. (2009). Principles of cancer therapy: Oncogene and non-oncogene addiction. *Cell*, 136: 823–837.

118. MacFarlane, L and Murphy. P. (2010) MicroRNA: Biogenesis, Function and Role in Cancer. *Current Genomics*, 11:7, 537-561.

119. Madu, C.O and Lu., Y. (2010). Novel diagnostic biomarkers for prostate cancer *Journal of Cancer*, 1:150-177.

120. Maniataki, E., Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fuelled by pre-miRNA. *Genes Development, 19*: 2979-2990.

121. Maret, W. (2014). Chapter 3: Potassium. In: Williams, R.J.P, editor. Binding, Transport and Storage of Metal Ions in Biological Cells, Cambridge University Press, 43-73.

122. Martinez-Ruiz, G., Maldonado, V., Ceballos-Cancino, G., Grajeda, J.P.R., Melendez-Zajgla, J. (2008). Role of Smac/DIABLO in cancer progression. *Journal of Experimental & Clinical Cancer Research* 27:48-52.

123. Mayeux. R., 2004, Biomarkers: Potential Uses and Limitations, Journal of Neurotherapeutics, 1(2):182-188.

124. McIlwain. D.R., Berger, T., Mak, T.W., (2013). Caspase Functions in cell death and disease. *Cold Spring Harbour Perspectives in Biology*, 7:11; 1-28.

125. Mehta, S., Shelling, A., Muthukaruppan, A., Lasham, A., Blenkiron, C., Laking, G., Print, C. (2010). Predictive and prognostic molecular markers for cancer medicine. *Therapeutic Advanced Medical Oncology*, 2(2): 125–148.

126. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, 2004, *15*, 185-197.

127. Melo, C.A., Melo S.A (2014). Biogenesis and Physiology of MicroRNAs. Non-coding RNAs and cancer. Edited by Muller Fabbri. Springer, New York, Heidelberg Dordrecht Londt.

128. Menendez, J.A., and Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer* 7:763-777.

129. Michael. A., Bajracharya, S.D., Yuen, P.S., Zhou, H., Star, R.A., Illei, G.G., and Alevizos, I. (2010). *Exosomes from human saliva as a source of microRNA biomarkers. Journal of Oral Diseases*, 16(1):34-8.

130. Michelsen, K.S., Thomas, L. S., Taylor, D.K., Yu, Q.T., Mei, L., Landers, C.J., Derkowski, C., McGovern, D.P.B., Rotter, J.I., Targan, S.R. (2009). IBD-Associated TL1A Gene (TNFSF15) Haplotypes Determine Increased Expression of TL1A Protein. *Public Library of Science Journal (PLOS)*, 10:161-170.

131. Min, Q., Ren, C.H., Sun, Y.H. (2014). Current early diagnostic biomarkers of prostate cancer. *Asian Journal of Andrology*, 16(4):549-54.

132. Misha, K., Adamusiak, T., Burdett, T., Culhane, A., Farn, A., Filippov, A. *et al*., (2012). Gene Expression Atlas update—a value-added database of microarray and sequencing-based functional genomics experiments, Nucleic Acid Research, 40(1):1077-1081.

133. Moldovan, L., Batte, K.E., Trgovcich, J., Wisler, J., Marsh, K.B., Piper, M. (2014). Methodological challenges in utilizing miRNAs as circulating biomarkers. *Journal of Cellular and Molecular Medicine*, 18(3):371-90.

134. Montagne, M., Naud, J.F., Lavigne. P. (2008). Elucidation of the structural determinants responsible for the specific formation of hetero-dimeric Mxd1/Max b-HLH-LZ and its binding to E-box sequences. *Journal of Molecular Biology*. 376:141-52.

135. Moreaux, J., Schoenhals, M., Kassambara, A., De Vos, J., Hose, D., and Klein, B. (2009). Embryonic stem cell markers expression in cancers. *Biochemical Biophysical Research Community* 29;383(2): 157-62.

136. Munker, R., and Calin, G.A. (2011). MicroRNA profiling in cancer. *Journal of Clinical Science*, 121:141–158.

137. Murphy D. (2002). Gene expression studies using microarrays: principles,

138. Nakanishi, T., Ross, D.D., Mitsuoka, K., (2010). Methods to evaluate transporter activity in cancer. *Methods in Molecular Biology*, 637:105-20.

139. National Collaborating Centre for Cancer. Prostate cancer: diagnosis and treatment. (2008). www.nice.org.uk/nicemedia/live/11924/39687/39687.pdf. Accessed May 18, 2015.

140. Negrini, S., Gorgoulis, V.G., Halazonetis, T.D. (2010). Genomic instability- an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, 11:220–228.

141. Negrini, S., Gorgoulis, V.G., Halazonetis, T.D. (2010). Genomic instability- an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, 11:220–228.

142. Nguyen, D.P., Li, J., Tewari, A.K. (2014). Inflammation and prostate cancer: the role of interleukin 6 (IL-6). *BJU International*, 113 (6): 986-92.

143. Oesterling, J.E., and M.A. Moyad. (1997). The ABCs of Prostate Cancer: The book that could save your life, Lantham, Md: Madison Books.

144. Ozsolak F *et al*., 2008. Chromatin structure analyses identify miRNA promoters. *Genes Development*, 22 (22):3172–83.

145. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, K., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., Hatzigeorgiou, A.G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research*, 41:180-188.

146. Parker R, Sheth U. P. (2007). Bodies and the control of mRNA translation and degradation. *Molecular Cell*, 25 (5):635-46.

147. Parkin. D. M., J. Ferlay, M. Hamdi-Ch´erif et al., 2003. Cancer in Africa—

148. Peltier, H. and Latham, G. (2008). Normalization of microRNA expression levels in quantitative RT-PCR assays: Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA*, 14(5): 844 – 852.

149. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9).

150. Pfaffl, M.W. (2004). Quantification strategies in Real Time PCR. In A-Z of Quantitative PCR. Ed S.A Bustin. International University Line. California, 87-112.

151. Pfaffl, M.W., Horgan, G.W., Dempfle, L. (2002). Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, 30(9):1-10.

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

153. Raza, K. (2012). Application of Data Mining in Bioinformatics. *Indian Journal of Computer Science and Engineering* 1(2):114-118.

154. Raza. K. (2012). Application of data mining in bioinformatics. *Indian Journal of Computer Science and Engineering*, 1(2):114-118.

155. Rebbeck. T. R., S. S. Devesa, B.L. Chang *et al*., (2013). "Global patterns of prostate cancer incidence, aggressiveness, and mortality in men of African descent," *Prostate Cancer*.

156. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Research*, 14(10):1902–10.

157. Roncaglia, P., Martone, M.E., Hill, D.P., Berardini, T.Z., Foulger, R.E., Imam, F.T., Drabkin, H., Mungall, C.J., Lomax, J. (2013). The Gene Ontology (GO)

Cellular Component Ontology: integration with SAO (Subcellular Anatomy Ontology) and other recent developments. *Journal of Biomedical Semantics*, 4(1): 1480-4.

158. Savli, H., Szendröi, A., Romics, I., Nagy, B. (2008). Gene network and canonical pathway analysis in prostate cancer: a microarray study. *Experimental and Molecular Medicine* 40 (2): 176–185.

159. Sawyers, C.L. (2008). The cancer biomarker problem. *Nature*, 452:548–52.

160. Selvaraj, S., Natarajan, J. (2011). Microarray Data Analysis and Mining Tools. *Nature*, 458:570–75.

161. Shin C, Nam J, Farh KK, Chiang HR, Shkumatava A, Bartel DP. (2010). Expanding the microRNA targeting code: functional sites with centered pairing. Molecular Cell, 38:789-802.

162. Siegel, R, L, (2015). Cancer Statistics 2015. American Cancer Society. Available at http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf

163. Sita-Lumsden, A., Dart, D.A., Waxman, J., Bevan, C.L. (2013). Circulating microRNAs as potential new biomarkers for prostate cancer. *British Journal of Cancer*, 108:1925-30.

164. Situ B, C, L. (2015). Cancer Facts and Figures 2015. American Cancer Society. Available http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2015/index

165. Su, Z., Yang, Z., Xu, Y., Chen, Y., Yu. Q., (2015). Apoptosis, autophagy, necroptosis, and cancer metastasis. *Journal of Molecular Cancer*, 14:48.

166. Tan, J., Yu, C.Y., Wang, C.H., Chen, H.Y., Guan, J., Chen, X.Y., Fang, J.Y. (2015). Genetic variants in the inositol phosphate metabolism pathway and risk of different types of cancer. *Scientific Reports* 5:8473-77.

167. Taylor. J. A., Albertsen PC. Benign and malignant disease of the prostate. In: Rosenthal RA, Zenilman ME, Katlic MR. (2011). *Principles and Practice of Geriatric Surgery*. 2nd Ed. New York, NY: *Springer Science and Business Media*, 4:1069-1082.

168. Theodorescu. D. Prostate cancer, clinical oncology. (2001). In: Schwab M, 5th Ed. *Encyclopedic Reference of Cancer*. Berlin, Germany: Springer-Verlag, 23:720-727.

169. Tollefson, M.K. (2008). Focus on PSA Screening for Prostate Cancer. *Clinical Update Mayo Clinic*, 28.

170. Tomaić, V., Gardiol, D., Massimi, P., Ozbun, M., Myers, M., Banks, L. (2008). Human and primate tumour viruses use PDZ binding as an evolutionarily conserved mechanism of targeting cell polarity regulators. *Oncogene* 28:1-8

171. Torre. L, Siegel. R, Jemal. A. (2012). Global Cancer Facts & Figures. *American Cancer Society*.

172. Triboulet, R *et al*., (2009). Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA*, 15(6):1005–11.

173. Triboulet, R., Chang, H.M., Lapierre, R.J., and Gregory, R.I. (2009). Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA*, 15:1005-1011.

174. van Peer, G., Mestdagh, P., Vandesompele, J. (2012). Accurate qRT-PCR gene expression analysis on cell culture lysates. *Scientific Reports* 2: 222-28.

175. Vaz, C., Hafiz, M.A., Bharti, R., Pandey, P., Kumar, L., Kulshreshtha, R., and Bhattacharya, A. (2013). Analysis of the microRNA transcriptome and expression of different isomiRs in human peripheral blood mononuclear cells. *BioMed Central Research Notes*, 6:390-95.

176. Velonas, V.M., Woo, H.H., dos Remedios, C.G., Assinder, S.J. (2013). Current Status of Biomarkers for Prostate Cancer. *International Journal of Molecular Science*, 14(6):11034-60.

177. Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos. I.L., Maniou, S., Karathanou, K., Kalfakakou, D., Fevgas, A., Dalamagas T., Hatzigeorgiou, A.G. (2014). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research* 42:153-159.

178. Wallace, D.M., Chisholm, G.D., Hendry, W.F. (1975). T.N.M. classification for urological tumours (U.I.C.C.) - 1974. *British Journal of Urology*, 47 (1):1-12.

179. Watanabe, S., Ueda, Y., Akaboshi, S., Hino, Y., Sekita, Y., Nakao, M. (2009).

180. Weinberg. R.A., and Hanahan.D. (2000). Hallmarks of Cancer. *Cell*. 100 (1):57–70.

181. Weinberg. R.A., and Hanahan.D. (2011). Hallmarks of Cancer: The Next Generation. *Cell*. 144(5):646–74.

182. Wong, M.L., Medrano, J.F. (2005). Real-time PCR for mRNA quantitation. *BioTechniques*, 39(1):1-11.

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

184. Worringer, A.L., Rand, A.L., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., Yamanaka, S. (2014). The ***let-7***/LIN-41 Pathway Regulates Reprogramming to Human Induced Pluripotent Stem Cells by Controlling Expression of Pro-differentiation Genes. *Cell Stem Cell*, 14, (1), 40-52.

185. Worringer, A.L., Rand, A.L., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., Yamanaka, S. (2014). The ***let-7***/LIN-41 Pathway Regulates Reprogramming to Human Induced Pluripotent Stem Cells by Controlling Expression of Pro-differentiation Genes. *Cell Stem Cell*, 14(1)40-52.

186. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X. and Li, T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*, 37:105–110.

187. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*, 37:105–110.

188. Yaman, F., Kovancilar, M., Dizdar, Y., Darendeliler, E., Holdenrieder, S., Dalay, N. *et al*., (2011). Investigation of miR-21, miR-141, and miR-221 in blood circulation of patients with prostate cancer. *Tumour Biology*,

189. Yan M, Huang HY, Wang T, Wan Y, Cui SD, Liu ZZ, (2012). Dysregulated Expression of Dicer and Drosha in Breast Cancer. *Pathological Oncology Research*, 18(2):343-8.

190. Yan M, Huang HY, Wang T, Wan Y, Cui SD, Liu ZZ, (2012). Dysregulated Expression of Dicer and Drosha in Breast Cancer. *Pathological Oncology Research*, 18(2):343-8.

191. Yang, J.S., and Lai, C.E. (2011). Alternative miRNA Biogenesis Pathways and the Interpretation of Core miRNA Pathway Mutants. *Molecular Cell*, 43(6): 892-903.

192. Ying, S. Y.; Lin, S. L. Intron-derived microRNAs--fine tuning of gene functions. *Gene*, 2004, *342*, 25-28.

193. Yu X, Lin J, Zack DJ, Qian J. (2006). Computational analysis of tissue specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Research*, 34(17):4925-4936.

194. Yu, X., Lin, J., Zack, D.J., Qian, J. (2007). Identification of tissue-specific cis-regulatory modules based on interactions between transcription factors. *BMC Bioinformatics*, 8(1):437.

195. Yuan, J.S., Reed, A., Chen, F., Stewart Jr, C.N. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7(85):1-1.

196. Zeegers, M. P., Kiemeney, L. A., Nieder, A. M., H Ostrer. (2004). "How strong is the association between CAG and GGN repeat length polymorphisms in the androgen receptor gene and prostate cancer risk?" Cancer Epidemiology Biomarkers and Prevention, 13 (11):1765–1771.

197. Zhang B, Pan X, Cobb GP, Anderson TA. (2007) MicroRNAs as oncogenes and tumour suppressors. *Developmental Biology 302*(1):1–12.

198. Zhang, HL, Yang LF, Zhu Y, Yao XD, Zhang SL, Dai B (2011). Serum miRNA-21: elevated levels in patients with metastatic hormone-refractory prostate cancer and potential predictive factor for the efficacy of docetaxel-based chemotherapy. *Prostate* 71: 326-331.

199. Zhang, Y., Eberhard, D.A., Frantz, G.D., Dowd, P., Wu, T.D., Zhou, Y., Watanabe, C., Luoh, S.M, Polakis, P., Hillan, K.J., Wood, W.I., Zhang, Z.

(2004). GEPIS–quantitative gene expression profiling in normal and cancer tissues. *Bioinformatics*, 20(15):2390-8.

200. Zhang, Y., Luoh, S.M., Lawrence, S., Baertsch, R., Wood, W.I., Zhang, Z. (2007). GeneHub-GEPIS: digital expression profiling for normal and cancer tissues based on an integrated gene database. *Nucleic Acids Research* 35:152-158.