



UNIVERSITY of the
WESTERN CAPE



Analysis, expression profiling and characterization of hsa-miR-5698 target genes as putative dynamic network biomarkers for prostate cancer: a combined *in silico* and molecular approach.

Chipampe Patricia Lombe

Student Number: 3111206

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

Advisor: Dr Ashley Pretorius

Co-advisor: Professor Mervin Meyer

January, 2019



Keywords

MicroRNA

Triplex

DNA sequence motif

Prostate cancer

Transcription factor

Promoter

Co-expression

qPCR

Pearson correlation

De novo

Abstract

In 2018, the International Agency for Research on Cancer (IARC) estimated that prostate cancer (PCa) was the second leading cause of death in males worldwide. The number of deaths are expected to raise by 50 % in the next decade. This rise is attributed to the shortcomings of the current diagnostic, prognostic, and therapeutic biomarkers used in the management of the disease. Therefore, research into more sensitive, specific and effective biomarkers is a requirement. The use of biomarkers in PCa diagnosis and management takes advantage of the genetic alterations and abnormalities that characterise the disease. In this regard, a microRNA, hsa-miR-5698 was identified in a previous study as a differentiating biomarker between prostate adenocarcinoma and bone metastasis. Six putative translational targets (CDKN1A, CTNND1, FOXC1, LRP8, ELK1 and BIRC2) of this microRNA were discovered using *in silico* approaches.

The aim of this study was to analyse via expression profiling and characterization, the target genes of hsa-miR-5698 in order to determine their ability to act as putative dynamic network biomarkers for PCa. The study was conducted using a combined *in silico* and molecular approach. The *in silico* part of the study investigated the putative transcriptional effects of hsa-miR-5698 on the promoters of its translational targets, the correlation between hsa-miR-5698 and mRNA expression profiles as well as the co-expression analysis, pathway analysis and prognostic ability of the target genes. A number of computational software were employed for these purposes, including, R Studio, Trident algorithm, STRING, KEGG, MEME Suite, SurvExpress and ProGgene. The molecular part of the study involved expression profiling of the genes in two PCa cell line LNCaP and PC3 via qPCR.

The study on the putative transcriptional effects of hsa-miR-5698 indicated that the microRNA has binding sites in the promoters of CDKN1A and FOXC1. Thus, forming a triplex structure with the promoter. Correlation analysis indicated a negative correlation between the microRNA and all its targets in both of the cell lines used (LNCaP and PC3), with the exception being FOXC1, whose expression profile was positively correlated to that of hsa-miR-5698 in PC3.

A co-expression analysis was done to identify genes putatively co-expressed with the target genes. Five genes were identified namely, CDK2, CDK4, TP53, CCND1 and PCNA. Pathway analysis of the co-expression network genes (CDKN1A, CTNND1, FOXC1, LRP8, ELK1, BIRC2, CDK2, CDK4, TP53, CCND1 and PCNA) indicated that CDKN1A was involved in the Phosphatidylinositol-3 kinase/Protein kinase B (PI3K/Akt) pathway specifically functioning in cell growth and proliferation. The MEME Suite package was used to identify transcription factor binding sites in the promoters of the 11 genes in the bid to understand the regulatory networks the genes are involved in. Two statistically significant novel transcription binding motifs were identified in the promoter sequences of the co-expression network genes. A search for transcription factors (TFs) matching the sequences of the motifs in the JASPAR database yielded one top match for each motif, namely FOXD3 and TBX2 corresponding to motif 1 and 2 respectively. The regulatory networks built using the identified putative TFs, co-expressed genes and hsa-miR-5698 identified two FFLs involved in PCa progression from adenocarcinoma to bone metastasis namely, the microRNA-FFL and the TF-FFL (FOXD3-FFL, TBX2-FFL).

The investigation into the prognostic value of the co-expression genes in terms of survival, recurrence and metastasis was undertaken via two online survival analysis tools, SurvExpress and ProGene. The datasets used for survival analysis and recurrence were the PRAD - TCGA - Prostate adenocarcinoma dataset, the

Galsky - Prostate - GSE45705 and the Kollmeyer-Jenkins Prostate GSE10645-GPL5858 each curating 497, 61 and 596 samples respectively. These datasets were obtained from the SurvExpress database. The metastatic analysis was undertaken on the Lapointe Prostate PNAS dataset curating 28 samples. This dataset was available from ProGene database. Results from two datasets indicated that CDKN1A, BIRC2 and FOXC1 might serve as good prognosis markers for metastasis in PCa. With presence of CDKN1A and BIRC2 indicating good prognosis for survival and relapse free survival and presence of FOXC1 indicating poor prognosis for the same. An evaluation of the metastatic prognostic ability of the target genes in the Human Cancer Metastasis Database (HCMD) yielded similar results indicating over-expression of FOXC1 in metastasised tissue samples but a down-regulation of CDKN1A and BIRC2.

The qPCR analysis of the expression profiles of hsa-miR-5698 targets and their co-expressed genes found that there was an up-regulation of CDKN1A, BIRC2 and CDK4 in LNCaP, and their down-regulation in PC3. There was also a down-regulation of FOXC1 in LNCaP and an up-regulation in PC3. This study indicated significant differences in the expression profiles of hsa-miR-5698 target genes between the two prostate cancer cell lines, which correspond to two different stages of the disease. The genes CDKN1A, BIRC2 and FOXC1 are able to distinguish between LNCaP and PC3 cell lines. The results obtained from the study indicate these biomarkers together with the microRNA hsa-miR-5698 could serve as network biomarkers in the monitoring and management of prostate cancer patients. These findings will be further investigated in human prostate tissues to validate these data.

C.P Lombe

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

DECLARATION

I declare that “Analysis, expression profiling and characterization of hsa-miR-5698 target genes as putative dynamic network biomarkers for prostate cancer: a combined *in silico* and molecular approach.” is my own work that has not been submitted for any degree or examination at any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Chipampe Patricia Lombe

January 2019



Signed _____

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 aunts and my mothers, Esther Jordan and Christine Lombe, thank you for your encouragement and support in the past 20 years and even more these past three years. To Uncle Brian and Uncle John, thank you for your being my inspiration.

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 helping to make my studies at UWC possible by facilitating the DST/MINTEK bursary. Thank you to the members of the BRG, Brother Marius, Zahra, Mohammed, Olalekan and Adewale for your continuous encouragement. 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 Dobby, thank you. Kashikulu, I would not have survived the past year without you, thank you my friend.

List of Abbreviations

AR	Androgen Receptor
AREs	Androgen Receptor Elements
ATCC	American Type Culture Collection
BIRC2	Baculoviral IAP Repeat-Containing Protein 2
BLAT	BLAST-like Alignment Tool
CANSA	The Cancer Association of South Africa
CCND1	Cyclin D1
CDK2	Cyclin Dependent Kinase 2
CDK4	Cyclin Dependent Kinase 4
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
ChIP	Chromatin immunoprecipitation assay
Cp	Crossing Point
CTNND1	Cadherin-Associated Protein Delta 1
DMSO	Dimethyl Sulphoxide
DRE	Digital Rectal Exam
E2F1	Transcription Factor 1
EGFR	Epidermal Growth Factor Receptor
ELK1	ETS Domain-Containing Protein 1
EMSA	Electrophoresis Mobility Shift Assay

EMT	Epithelial–Mesenchymal Transition Pathway
ERG	Erythroblast Transformation-specific Regulator Gene
FBL	Feedback Loops
FBS	Fetal Bovine Serum
FFLs	Feed Forward Loops
FOXC1	Forkhead Box C1 Transcription Factor
FOXD3	Forkhead Box D3
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GEO	Gene Expression Omnibus
GO	Gene Ontology
GOMo	Gene Ontology for Motifs
HCMD	Human Cancer Metastasis Database
HPRD	Human Proteinpedia Database
HPRT-1	Hypoxanthine-Guanine Phosphoribosyltransferase
IARC	International Agency for Research on Cancer
ICGEB	International Centre for Genetic Engineering and Biotechnology
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LOE	Level of Evidence
LRP8	Lipoprotein Receptor-Related Protein 8
MAGIA	MicroRNA and Genes Integrated Analysis
MEME	Multiple Em for Motif Elicitation

MREs	MicroRNA-Recognition Elements
mRNA	Messenger RNA
PCa	Prostate Cancer
PCA3	Prostate cancer Antigen 3
PCNA	Proliferating Cell Nuclear Antigen
PI3K	Phosphatidylinositol-3 kinase
PIK3CA	phosphatidylinositol 3-kinase
PSA	Prostate Specific Antigen
PTEN	Phosphatase and Tensin Homolog
PTEN	Phosphatase and Tensin Homolog
qPCR	Quantitative Real Time Polymerase Chain Reaction
RNA	Ribonucleic Acid
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SMAD4	Signal Transduction Protein 4
SNP	Single Nucleotide Polymorphism
SPP1	Secreted Phosphoprotein 1
SRA	Sequence Read Archive
STRING	Search Tool for the Retrieval of Interacting Genes
TBX2	T-box Transcription Factor 2
TCGA	The Cancer Genome Atlas
TF/s	Transcription Factor/s

TFBSs	Transcription Factor Binding Sites
TGF- β 1	Transforming Growth Factor Beta 1
TiGER	Tissue-Specific Gene Expression and Regulation
TMPRSS2	Transmembrane protease serine 2:v- ets erythroblastosis virus E26 oncogene homolog
TNRC6	Trinucleotide Repeat Containing 6
TP53	Tumour 53
TSS/s	Transcription Start Site/s
USPSTF	United States Preventive Services Task Force
UTR	Untranslated region
WHO	World Health Organisation
ZNF263	Zinc Finger Protein 263

List of figures

Chapter 1

Figure 1.1: The leading cause of death by cancers worldwide (for males) in 2018, with prostate cancer accounting for 13.5 %.

Figure 1.2: Pathways of androgen biosynthesis in normal physiology and prostate cancer

Chapter 2

Figure 2.1 : Hoogsteen base pair between adenine and uracil (A).

Figure 2.2: Molecular model of hsa-miR-5698-CDKN1A DNA triplex.

Figure 2.3: Molecular model of hsa-miR-5698-FOXC1 DNA triplex.

Figure 2.4: Correlation analysis showing the strength of the relationship between hsa-miR-5698 and its targets in the PC3 and LNCaP prostate cancer cell lines.

Chapter 3

Figure 3.1: Feed Forward Loops classified by the master regulator. Blunt arrows with dot end represent transcriptional activation or repression, T-shaped arrow represents repression.

Figure 3.2: Feedback Loops. A is the single negative loop and B is the double negative loop. The sharp arrow indicates activation.

Figure 3.3: Protein co-regulation network from STRING and IntAct, visualized in Cytoscape. The microRNA gene targets are outlined in the black boxes and the co-regulated genes are outlined in blue.

Figure 3.4: The prostate cancer pathway from KEGG database. The pathway shows the position of three of the genes of interest in the pathway (red boxes).

Figure 3.5: Sequence logos of three identified motifs common among the microRNA target genes and their co-regulated genes.

Figure 3.6: Sequence logos of the reverse complements (antisense strand) of the three identified motifs common among the microRNA target genes and their co-regulated genes.

Figure 3.7: Location and p-value of the three identified *de novo* motifs on the promoters of the submitted genes.

Figure 3.8: Venn diagram of motif distribution on the gene promoters.

Figure 3.9: Optimal alignment of motif 1 and the logo of the query.

Figure 3.10: Optimal alignment of motif 2 and the logo of the query.

Figure 3.11: Optimal alignment of motif 3 and the logo of the query.

Figure 3.12: Integrative microRNA target genes and the DNA sequence motif 1 transcription factor FOXD3 regulatory networks in the prostate adenocarcinoma cell line LNCaP.

Figure 3.13: Integrative microRNA target genes and the DNA sequence motif transcription factor FOXD3 regulatory networks in the prostate metastatic cell line PC3.

Figure 3.14: Integrative microRNA target genes and the DNA sequence motif 2 transcription factor TBX2 regulatory networks in the prostate metastasis cell line PC3.

Figure 3.15: Integrative microRNA target genes and the DNA sequence motif 2 transcription factor TBX2 regulatory networks in the prostate adenocarcinoma cell line LNCaP

Chapter 4

Figure 4.1: Gene expression values for each of the genes in each cohort (High risk red and Low risk green) represented by box and whisker plots. Lapointe Prostate PNAS dataset.

Figure 4.2: Combined Kaplan Meier analysis for four candidate genes for prostate cancer metastasis in the SurvExpress database.

Figure 4.3: Heat map demonstrating expression of each of the candidate genes in each of the 29 samples used in the dataset.

Figure 4.4: Combined Kaplan Meier analysis plot all microRNA target genes for prostate cancer relapse free survival in the ProGgene server.

Figure 4.5: Combined Kaplan Meier analysis plot of all microRNA target genes for prostate cancer overall survival in the ProGgene server in the GSE16560 dataset.

Figure 4.6: Combined Kaplan Meier analysis plot of all microRNA target genes for prostate cancer overall survival in the ProGgene server.

Figure 4.7: Comparison of gene expression profiles in prostate normal tissue, prostate adenocarcinoma primary tumour and metastatic tumour from The Human Cancer Metastasis Database (HCMD).

Chapter 5

Figure 5.1: Relative expression ratio plot of the targets of hsa-miR-5698 targets and their co-expressed genes in the prostate cancer cell lines LNCaP and PC3.

List of tables

Chapter 1

Table 1.1: Levels of evidence for prognostic studies.

Chapter 2

Table 2.1: Putative interaction of hsa-miR-5698 and the promoters of its translational targets.

Table 2.2: hsa-miR-5698 and their corresponding genes show correlating expression profiles.

Chapter 3

Table 3.1: Position and size of motif 2 on the promoters of the corresponding genes.

Table 3.2: Top matched transcription profiles for each *de novo* motif.

Table 3.3: Representation of the total number of GO terms predicted per motif

Table 3.4: Summary of relationships among PCa-related genes, hsa-miR-5698, and putative TFs.

Chapter 5

Table 5.1: General reagents and suppliers

Table 5.2: Cell lines used to investigate gene expression in prostate cancer.

Table 5.3: Primer sequences for PCR amplification of cDNA

Table 5.4: Reagents and components for cDNA synthesis

Table 5.5: Reagents for a standard qPCR reaction

Table 5.6: Cycling Protocol for the qPCR

Table 5.7: Specific LightCycler® 480 parameters for the qPCR

List of equations

Chapter 2

2.1 Pearson Correlation analysis

Appendix

List of tables

Table 7.1: Cis-Regulatory Module (CRM) detections in prostate cancer.

List of figures

Figure 7.1: Kaplan Meier analysis plots for CDKN1A (1) and CTNND1 (2) for prostate cancer relapse free survival in the ProGgene server in the GSE70769 dataset.

Figure 7.2: Kaplan Meier analysis plots for ELK1 (1) and BIRC2 (2) for prostate cancer relapse free survival in the ProGgene server in the GSE70769 dataset.

Figure 7.3: Kaplan Meier analysis plots for FOXC1 (1) and LRP8 (2) for prostate cancer relapse free survival in the ProGgene server in the GSE70769 dataset.

Figure 7.4: Kaplan Meier analysis plots for CDKN1A (1) and CTNND1 (2) for prostate cancer relapse free survival in the ProGgene server in the TCGA-PRAD dataset.

Figure 7.5: Kaplan Meier analysis plots for ELK1 (1) and BIRC2 (2) for prostate cancer relapse free survival in the ProGgene server in the TCGA-PRAD dataset.

Figure 7.6: Kaplan Meier analysis plots for FOXC1 (1) and LRP8 (2) for prostate cancer relapse free survival in the ProGgene server in the TCGA-PRAD dataset.

Table of Contents

Keywords	i
Abstract	ii
DECLARATION	v
DEDICATION	vi
Acknowledgements	vii
List of Abbreviations	viii
List of figures	xii
List of tables	xv
List of equations	xvi
Chapter 1	1
Literature Review	1
1.1 Prostate Cancer	1
1.2 Development of prostate cancer	3
1.3 Risk factors of PCa	3
1.3.1 Non-environmental risk factors	4
1.3.2 Age	4
1.3.3 Family history	5
1.3.4 Germline genetics of prostate cancer	7
1.3.5 Somatic molecular alterations in PCa	8
1.3.5.1 PTEN (Phosphatase and Tensin Homolog)	9
1.3.5.2 PIK3CA	10
1.3.5.3 TMPRSS2-ERG gene fusion	11
1.3.6 Race	13
1.3.7 Hormones as risk factors	14
1.3.8 Environmental risk factors of prostate cancer	17
1.4 Screening, diagnosis and prognosis of prostate cancer	18
1.4.1 Prostate specific Antigen	19
1.4.2 Digital rectal examination (DRE)	21
1.4.3 The prostate cancer antigen 3 gene (PCA3)	22

1.4.4 Prostate biopsy.....	23
1.4.5 Prostate cancer prognosis.....	24
1.4.6 The effect of current prognostic biomarkers on management of prostate cancer	27
1.4.6.1 Active surveillance	27
1.4.6.2 Radical prostatectomy	27
1.4.6.3 Adjuvant treatment	28
1.5 MicroRNAs.....	29
1.5.1 MicroRNAs and microRNA target genes as network biomarkers for PCa.....	30
1.6 Bioinformatics as a tool for the detection of novel biomarkers	32
1.7. Study rationale	33
1.7.1 Aims and Objectives.....	34
Aim.....	34
Objectives	34
1.8 References	36
Chapter 2	54
MicroRNA and Target Gene Correlation analysis.....	54
2. 1 Introduction.....	54
2.1.2 MicroRNAs form triplexes with double stranded DNA at sequence specific binding sites to alter gene expression	55
2.1.3 Pearson Correlation Analysis	57
2.2 Aims and objectives	59
2.2.1 Previous work.....	59
2.3 Methodology	61
2.3.1 Identification of putative transcriptional effects of hsa-miR-5698 on its translation targets' promoter sequences.	61
2.3.1.1 Extraction of gene promoters from the Ensembl database	61
2.3.1.2 MicroRNA triplex binding sites on promoters of target genes.....	61
2.3.2 Visualization of microRNA and promoter DNA triplex	62
2.3.3 Correlation of expression profiles between hsa-miR-5698 its mRNA targets in LNCaP and PC3 cell lines.	62
2.4 Results and discussion	63
2.4.1 hsa-miR-5698 as a possible transcriptional target of its translational target genes	63

2.4.2 Correlation between hsa-miR-5698 its targets expression profiles in LNCaP and PC3 cell lines.	68
.....	
2.5 Conclusion	71
2.6 References	74
Chapter 3	80
MicroRNA and Transcription Factor gene co-regulation	80
3.1.1 Background	80
3.1.2 Regulation of Gene Expression	81
3.1.2.1 Promoters	82
3.1.2.2 Transcription factors (TFs)	82
3.1.2.3 MicroRNAs	83
3.1.3 Transcription factor and microRNA co-regulatory networks	84
.....	85
3.1.4 DNA Sequence motifs (transcription factor binding sites)	86
3.1.5 Gene/protein interaction analysis using the STRING database	88
3.1.6 The IntAct database for co-expression analysis	89
3.1.7 The KEGG Pathway database	90
3.1.8 Ensembl	90
3.1.9 MEME (Multiple Em for Motif Elicitation)	91
3.1.9.1 TOMTOM	92
3.1.9.2 Gene Ontology for Motifs (GOMo)	92
3.1.10 Previous work	94
3.2 Methodology	96
3.2.1 (a) Co-expression analysis via STRING	96
3.2.1 (b) Co-expression analysis via IntAct	96
3.2.2 Extraction of gene promoters from the Ensembl database	97
3.2.3 <i>De novo</i> motif discovery using MEME (Multiple Em for Motif Elicitation) SUITE	98
3.2.4 Comparison of discovered motifs to elements of databases of known motifs using TOMTOM	98
3.2.5 Gene Ontology (GO) analysis for the discovered motifs using Gene Ontology for Motifs (GOMo)	99
3.2.6 Building the regulatory relationships among hsa-miR-5698, its target genes and putative transcription factors in PCa	99
3.3 Results and discussion	100

3.3.1 Co-expression analysis via STRING and IntAct	100
3.3.2 KEGG pathway analysis	103
3.3.3. <i>De novo</i> motif discovery	107
3.3.4 Comparing the discovered motifs to known transcription factor binding profiles in the JASPAR database using TOMTOM.....	115
3.3.5 Gene Ontology (GO) analysis via GOMo	121
3.3.6 MicroRNA and DNA sequence motifs co-regulation of gene expression.....	122
3.3.7 Conclusion.....	129
3.4 References	132
Chapter 4	144
4.1 Background.....	144
4.1.1 Prognostic determinants in Prostate cancer (PCa)	145
4.1.2 Clinical factors	146
4.1.2.1 Serum Markers - Prostate Specific Antigen (PSA).....	146
4.1.2.2. Gleason score	147
4.1.3. Molecular factors	148
4.1.3.1 Tumours markers	148
(i) Tumour markers involved in evasion of apoptosis.....	149
(ii) Self-sufficiency in growth signals	149
(iii) Signature markers of lethal prostate cancer	150
4.1.3.2 Urine markers.....	151
4.1.3.3 The role of bioinformatics in biomarker discovery.....	152
4.2.1 ProGgene	153
4.2.2 SurvExpress®	153
4.2.3 HCMDB: the human cancer metastasis database	154
4.3 Aims and Objectives.....	155
4.4 Methodology	156
4.4.1. Log-rank survival, recurrence and metastasis analysis of the prognostic value of the target genes via SurvExpress	156
4.4.2 Log-rank relapse free survival and death analysis of the prognostic value of the target genes via ProGgene	157

4.4.3 Evaluation of the metastatic prognostic ability of the target genes using the Human Cancer Metastasis Database (HCMD)	157
4.5 Results and discussion	158
4.5.1 Log-rank metastatic analysis of the prognostic value of the target genes in prostate cancer via SurvExpress	158
4.5.2 Log-rank relapse free survival (recurrence) analysis of the prognostic value of the target genes in prostate cancer via ProGgene.	164
4.5.3 Log-rank overall survival analysis of the prognostic value of the target genes in prostate cancer via ProGgene.....	167
4.5.4 Metastatic prognostic ability of the target genes using the Human Cancer Metastasis Database (HCMD)	169
4.6 Conclusion	172
4.7 References	174
Chapter 5	183
Molecular Validation of microRNA-target genes as Putative Biomarkers for the Detection and progression of Prostate Cancer.	183
5.1 Introduction.....	183
5.1.1 Quantitative real-time PCR (qPCR).....	184
5.1.3 Aims and objectives	186
5.2 Materials and methods	187
5.2.1. Cell culture.....	188
5.2.2. Start-up of cell culture from frozen cells.....	188
5.2.3. Maintaining the cell lines.....	189
5.2.4. Sub-cultivation and trypsinization for mRNA extraction.....	189
5.2.5. Extraction of RNA	190
5.2.6. Primer design	191
5.2.7. Reverse transcription of mRNA to cDNA.....	192
5.2.8 Analysis of gene expression profiles of the genes in cancer and control cell lines using qPCR	194
5.3 Results and Discussion	197
5.4 Conclusion	201
5.4 References	203
Chapter 6	210

General Discussion and Conclusion	210
6.1 General discussion	210
6.2 Conclusion	218
6.3 Future Work	220
6. 4 References	221
Appendix A	225

Chapter 1

Literature Review

1.1 Prostate Cancer

Cancer is currently one of the greatest health concerns worldwide, with 17 million new cases of cancer reported in 2018 (IARC, 2018; WHO, 2018). After Lung cancer, prostate cancer is the second leading cause of death in men globally (Figure 1.1) (IARC, 2018). Key statistical reports for 2018 by the World Health Organisation (WHO) reported that 70 % of deaths from prostate cancer occur in low and middle income countries, specifically in Africa, Asia and Latin America (WHO, 2018). Some of the reasons given for the large percentage of deaths in these populations include late stage presentation of the disease, which is mainly attributed to the lack of accessible diagnostics, pathology services and treatment facilities, available to citizens of countries in these continents, especially in the public health institutions. (IARC, 2018; WHO, 2018).

The Cancer Association of South Africa (CANSA) reports that one in six males and one in seven females have some type of cancer with prostate cancer (PCa) being most prevalent in the former and breast cancer in the latter (Herbst and Joubert 2017).

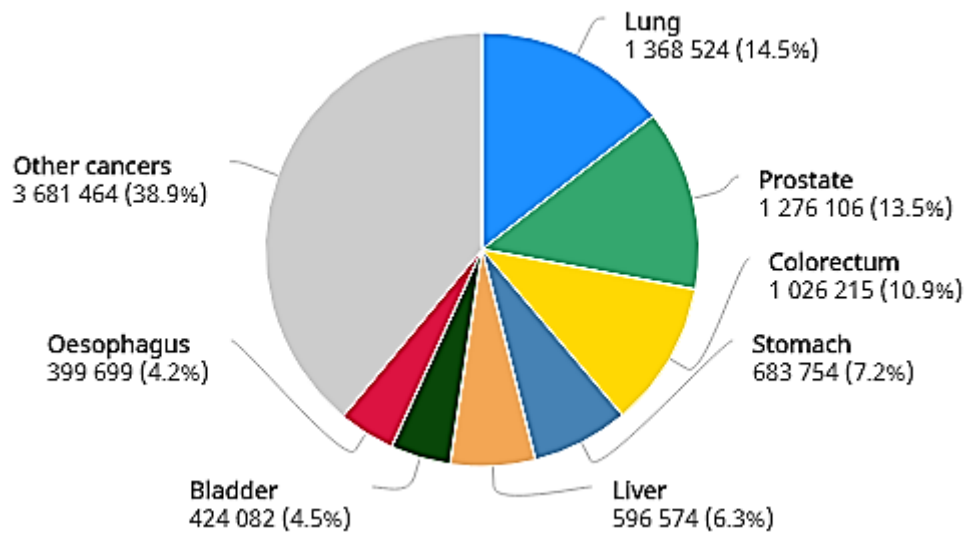


Figure 1.1: The leading cause of death by cancers worldwide (for males) in 2018, with prostate cancer accounting for 13.5 %. (Adapted from IARC, 2018).

However, a study by Roux *et al.*, 2015 questions such figures, stating that they could be an underestimation of PCa cases in South Africa. In their study on PCa among different racial groups in the Western Cape, Heyns *et al.*, 2011, attributed that such 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 diagnostic methods such as the digital rectal examinations.

1.2 Development of prostate cancer

Two factors contributing to the development and progression of prostate cancer are hereditary (genetic) and/or environmental factors (Migliore and Coppedè, 2002; Parsa, 2012). It has been proposed however, that the environment is the dominant determinate in the variability of cancer from population to population as well as from country to country.

This is because, while there are a number of disease causing alleles distributed unequally in populations, they cannot explain the dramatically different incidence rates of various cancers throughout the world (Weinberg 2013; Crocetti *et al.*, 2017). This has been demonstrated by measuring cancer rates in migrant populations. The rate of stomach cancer in Japanese populations is 4 to 8 times higher than that of Americans. A case study was conducted on Japanese migrants settling in the United States, within a generation, their offspring exhibit stomach cancer rates that are comparable to that of the surrounding population (Peto, 2010). Thus, for the great majority of cancers, disease risk therefore seems to be “environmental,” where this term is understood to include both physical environment and lifestyle.

1.3 Risk factors of PCa

There is no single factor considered to be solely responsible for the development of PCa. However, there are a number of risk factors that have been identified which are associated with its pathogenesis and development. Evidence for these risk factors is from both observational studies as well as randomised controlled trials (Fan *et al.*, 2018; Shakil *et al.*, 2018). As such, prostate cancer risk factors

may be presented as, (i) non-environmental risk factors and (ii) environmental risk factors.

1.3.1 Non-environmental risk factors

These are classified as those factors that have a strong association with PCa risk and are currently beyond human or medical intervention such as age, family history of PCa, genetics, race and hormonal levels (Godtman *et al.*, 2016; Roberts *et al.*, 2018).

1.3.2 Age

One of the strongest risk factors of prostate cancer is age. Studies have shown that occurrence of PCa in individuals younger than 50 years old is uncommon. With less than 0.1 % of diagnosed patients falling under the age of 50 and over 75 % of diagnosed patients falling above the age of 65 (Godtman *et al.*, 2016; IARC, 2018). Recently, the cumulative risk of PCa at the age of 74 has been calculated to be 12.8 % worldwide (Godtman *et al.*, 2016; Cancer Research UK, 2018). However, there has been a recent age migration of the disease over the last decade. This has been brought about by the availability of screening methodologies such as the prostate specific antigen (PSA) biomarker test (Godtman *et al.*, 2016; Roberts *et al.*, 2018).

Thus, a relatively larger proportion of men are now diagnosed at an earlier age, which has also led to earlier staging and grading of the disease (Godtman *et al.*, 2016). However, the benefits of this stage and age shift on the quality of life for

PCa patients remains unclear and is widely debated (Godtman *et al.*, 2016; Braunhut *et al.*, 2018; Cancer Research UK, 2018). Further research into survivorship experience of clinically presented and screen detected patients may provide valuable information on harms and benefits of early diagnosis.

1.3.3 Family history

A family history has been firmly established as a risk factor for prostate cancer (Drake *et al.*, 2008; Rudichuk *et al.*, 2016). When examining the family history of PCa, two classifications can be made for the disease, namely familial and hereditary (Rudichuk *et al.*, 2016). Familial PCa is marked by one or more than one first degree relative being affected by the disease (Vertosick *et al.*, 2014; Rudichuk *et al.*, 2016). Hereditary PCa shows a pattern of cancer distribution resulting from Mendelian inheritance of a susceptibility gene (Vertosick *et al.*, 2014).

There have been a number of studies that have shown an increased risk of PCa for sons and brothers of men with the disease. The relative risk of an individual to develop PCa increases significantly according to the number of individuals affected in family, their relationship with the index case/s as well as the age at which they developed the disease (Liss *et al.*, 2014; Selkirk *et al.*, 2015).

Studies have shown that if a brother or father of an individual had PCa, the relative risk of PCa in that individual is doubled with risk of 15 % (Drake *et al.*, 2008; Giri and Beebe-Dimmer, 2016). If the aforementioned relative had the disease before the age of 60, the relative risk increases three fold (Giri and

Beebe-Dimmer, 2016). As a result, individuals with hereditary predispositions to PCa are generally diagnosed as much as six to seven years earlier than those with sporadic cancer (Giri and Beebe-Dimmer, 2016). However, there are no differences between the clinical features of patients with hereditary and sporadic PCa in terms of tumour grade and pathological stage at diagnosis (Selkirk *et al.*, 2015; Giri and Beebe-Dimmer, 2016).

A comparison of the survival of sporadic and familial PCa patients shows that there is no significant difference observed between the two (Roehl *et al.*, 2006; Stewart *et al.*, 2017). However, such a comparison can be difficult to undertake, this is because of the earlier stage of diagnosis, differentiation and localisation of disease among hereditary cancers. Some studies show that these characteristics lead to longer survival rates (Lythgoe *et al.*, 2016; Blackwelder and Chessman, 2018). In addition, PCa is a complex disease with its initiation arising from interaction between genetic and non-genetic factors. Perhaps the factors of how it arises may have an impact on survival rates.

Even though prostate cancer exhibits the highest reported heritability of any major cancer, the ability to define hereditary prostate cancer genes has been limited (Hjelmborg *et al.*, 2014; Mucci *et al.*, 2016; Rebbeck, 2017). However, some genes responsible for hereditary prostate cancer that have been identified mainly in European populations. These include HPC1 (1q24-25), PCAP (1q42-43) (Berry *et al.*, 2000), HOXB13 (Breyer *et al.*, 2012; Xu *et al.*, 2013) and HPC20 (Berry *et al.*, 2000; Rebbeck, 2017). These data remain largely uncorroborated in African populations. As a result, of this, there is a lack of genetic testing for hereditary PCa and subsequent recommendations for risk reduction in clinical practice in many African countries (Rebbeck, 2017). Several studies have been conducted on African populations in the Americas, but how geographic location

and differences in gene pools affects the disease incidence has not been extensively detailed (Rebbeck, 2017; Nettey *et al.*, 2018).

1.3.4 Germline genetics of prostate cancer

Prostate cancer has strong heritable components and deleterious germline variants in certain genes can increase the risk of the disease (Patrick *et al.*, 2017). Genome-wide association studies have identified over 100 single nucleotide polymorphisms (SNPs) associated with prostate cancer risk (Spencer *et al.* 2009; Goh *et al.* 2012; Dias *et al.*, 2017). Single nucleotide polymorphisms in the BRCA1, BRCA2, MMR, HOXB13, CHEK2, and NBS1 genes confer moderate risks with some leading to aggressiveness of the disease (Mikropoulos *et al.*, 2014; Dias *et al.*, 2017).

The gene HOXB13 is currently of interest in prostate cancer progression. This gene functions in segmentation during embryonic development. It interacts with the androgen receptor in both the normal prostate as well as in PCa. Thus, it is an important regulator of cellular response to androgens (Kim *et al.*, 2010; Dias *et al.*, 2017). A G84E SNP in this gene has been recorded by a number of studies and has been associated with increased risk of hereditary PCa as well as progression and aggression as the mutation encourages proliferation of androgen independent cells (Dias *et al.*, 2017; FitzGerald *et al.*, 2017; Johng *et al.*, 2017).

Another gene that has been associated with germline inheritance of prostate cancer is CHEK2; this gene encodes a G2 checkpoint kinase involved in DNA repair leading to cell-cycle arrest, activation, apoptosis and cell death. The

mutation of this gene has been linked to high risk of breast cancer as well as prostate cancer (Wang *et al.*, 2015). However, a recent study found that unlike the HOXB13, CHEK2 could not differentiate between the risk of lethal and indolent disease (Zheng *et al.*, 2017). Thus, no link to disease progression and aggressiveness is currently known.

Germline variants can influence diverse treatment modalities and thus making use of genetic biomarkers to guide treatment decisions is highly beneficial to clinical practice. A study by Kearns *et al.*, 2016 on a cohort of surgically treated PCa patients showed that an SNP, rs11568818 located on the MMP7 gene is associated with pathological upgrading of the disease (Kearns *et al.*, 2016). This finding has also been confirmed in a second cohort of patients on active surveillance (Kearns *et al.*, 2016) as well as in another study on a different cohort with the same conditions (Cooper *et al.*, 2017). Statistics show that approximately 30 % of patients on PCa active surveillance move on to definitive treatment in the first five years of perception of the cancer (Dias *et al.*, 2017), because of this, precise genetic information would be of great value in informing treatment stratification.

1.3.5 Somatic molecular alterations in PCa

Prostate tumour development and progression involves alterations in numerous genetic pathways. As the understanding of these pathways and molecular alterations has evolved, the markers involved have also been elucidated. Several markers have been identified for their roles in the prediction of disease outcome as well as therapeutic targets. (Nelson *et al.*, 2009; Netto *et al.*, 2017). Some of

the specific markers identified include the tumour suppressor genes (p53, p21, Rb and PTEN), the oncogenes (c-myc, HER2 and Bcl2), the apoptosis regulators (survivin and TGF β), the prostate tissue lineage specific markers (PSA and PCA3), and the adhesion molecules (CD44 and PI3K/akt/mTOR pathway members (Netto, 2015; Engelstein, 2017; Neto *et al.*, 2017). Somatic mutations of some of the markers above has been related to high incidences of metastatic PCa and variations in treatment response.

There are a number of candidate genes such as AR, TP53, PIK3CA, CHEK2, PTEN, KIT, c-myc, FOXA1 and ZFH3 that are prone to somatic mutations in the development and progression of prostate cancer (Tapia-Laliena *et al.*, 2014; Alvarez-Cubero *et al.*, 2018; Martinez-Gonzalez *et al.*, 2018). However, there is very little information available for somatic mutations involved in tumour development at the metastatic stage (Alvarez-Cubero *et al.*, 2017). Perhaps this is due to the difficulty in obtaining enough quality tissue to undertake the studies since bone is the primary site of metastasis in PCa (Mehra *et al.*, 2011). Nevertheless, a recent study found that a substitution – missense mutation in PIK3CA at position 1047, H \rightarrow R results in poor prognosis for the disease and could lead to metastasis (Pearson *et al.*, 2018).

1.3.5.1 PTEN (Phosphatase and Tensin Homolog)

The gene PTEN is a tumour suppressor gene found on chromosome 10q23.3. It encodes a 403 amino acid protein involved in the cell cycle. It regulates cell cycle progression by maintaining the G2/S phases of the cell cycle checkpoint (Mulholland *et al.*, 2012). There are a great number of studies focused on the

relationship of PCa and mutations related to it in the disease. A recent study found that a homozygous loss of function mutation of the gene is associated with a higher Gleason score, regional pathological staging and recurrence of disease after radical therapy (Geybels *et al.*, 2017; Jamaspishvili *et al.*, 2018). PTEN loss is also associated with suppression of androgen receptor (AR) transcriptional output. Phosphoinositide 3-kinase (PI3K) inhibitors are known to activate the AR signalling, perhaps a combination of therapies targeting the PI3K and AR signalling pathways may enhance treatment efficacy.

1.3.5.2 PIK3CA

The gene PIK3CA is a gene that codes for a protein subunit that forms part of the phosphatidylinositol 3-kinase. It is located on chromosome 3q26.32 (Alvarez-Cubero *et al.*, 2018). In many cancers including breast and colorectal cancers, PIK3CA undergoes a mutation in the p110 alpha subunit of the PI3 kinase enzyme. This mutation rises late in tumorigenesis, just before or concurrent with invasion and migration (Samuels and Waldman, 2010; Alvarez-Cubero *et al.*, 2018). Martinez-Gonzalez *et al.*, 2018 evaluated 125 prostate biopsies for genes undergoing somatic mutations. They found that in 50 % of their subjects, PIK3CA underwent either of two substitutions (3129G > T and 3139C > T) in the aforementioned subunit that results in a nonsense mutation (Martinez-Gonzalez *et al.*, 2018). The altering of the PI3K enzymes inhibits its entrance into the PI3K/Akt/mTOR pathway and is linked to the survival and metastasis of PCa cells, the development of drug resistance as well as development of castration resistant PCa (Robinson *et al.*, 2015; Tang and Ling, 2014).

1.3.5.3 TMPRSS2-ERG gene fusion

The transmembrane protease serine 2:v-ets erythroblastosis virus E26 oncogene homolog (TMPRSS2-ERG) gene fusion is a common occurrence in 50 % of prostate cancers (Tomlins *et al.*, 2008; Wang *et al.*, 2017). However, its functional role is not fully understood. There are a number of studies that have been conducted to elucidate the function of this gene fusion in human prostate cancers.

The TMPRSS2 gene encodes a protein that belongs to the serine protease family. Thus, its protein product is a transmembrane receptor (Ko *et al.*, 2015). The gene is located on the human chromosome 21. A significant feature of this gene that begins to highlight its importance in prostate cancer is that it has several androgen receptor elements (AREs) located upstream of its transcription start site (Lucas *et al.*, 2014; Shen *et al.*, 2017). It has been demonstrated that TMPRSS2 activates the protease activated receptor 2 (PAR-2), which is a G-protein coupled receptor. This activation causes an up-regulation of an enzyme called metalloproteinase-2 (MMP-2), which is a key protease in tumour metastasis (Lucas *et al.*, 2014; Shen *et al.*, 2017). Additionally, this gene is up-regulated by androgenic hormones in prostate cancer cells and down-regulated in androgen-independent prostate cancer tissue (Ko *et al.*, 2015). Therefore it could be a distinguishing clinical marker for cancer progression and perhaps treatment.

The ERG gene encodes the erythroblast transformation-specific (ETS) protein which is a member of the transcription factor family. This transcription factor is one of a number of key regulators of cell proliferation, differentiation, angiogenesis, inflammation, and apoptosis (Lee *et al.*, 2018).

In a recent study, Wang and co-workers (2017), found that the TMPRSS2 and ERG genes were mostly altered in prostate cancer, and the most frequent alteration was a translocation of TMPRSS2 resulting in the gene fusion (Wang *et al.*, 2017). Additionally, examination of four independent prostate cancer datasets from Oncomine and cBioportal demonstrated that the ERG gene was significantly upregulated in prostate clinical samples compared with the normal prostate gland fusion (Wang *et al.*, 2017). Another study by Hägglöf *et al.*, 2014 knocked down ERG and found a cell-cycle arrest at G₀/G₁ phase, and consequently, a reduction in cell proliferation (Hägglöf *et al.*, 2014).

This gene fusion could thus serve as an important marker in clinical and prognostic monitoring. In this regard, some studies have been conducted to characterise the difference between TMPRSS2-ERG and non-TMPRSS2-ERG fusion PCa patients and the possible implications. Some studies conducted on this fusion gene have identified a statistically significant association between the TMPRSS2-ERG fusion and prostate cancer specific death. Thus, these results suggest that prostate cancer with TMPRSS2-ERG fusion is a distinct molecular subtype. However, there exists great diversity in the precise structure of the TMPRSS2-ERG hybrid transcripts in human prostate (Demichelis *et al.*, 2007; The Cancer Genome Atlas Research Network, 2015; Arora and Barbieri, 2018). Perhaps PCa sub-typing and specific death is as a result of a particular subtype. Additionally, the TMPRSS2-ERG fusion has been linked to a more aggressive tumour stage as well as a PCa phenotype (Hägglöf *et al.*, 2014; Kulda *et al.*, 2016; Sanda *et al.*, 2017).

There is increasing evidence of the importance of the TMPRSS2-ERG gene in prostate cancer prognostics. However, for the percentage of patients who do not present with this gene fusion, other biomarkers need to be investigated.

1.3.6 Race

Several studies have linked a higher risk of PCa development in black men when compared to Caucasians and Asian men. With black men being 1.5 to 2 times more at risk of developing the disease than white men. However, white men have a higher risk of the disease when compared to their Asian counterparts (Rebbeck, 2016; Nettey *et al.*, 2018; Li *et al.*, 2018). The differences in the incidence of PCa raise questions as to whether they are as a result of genetic predisposition of certain races to develop the disease or a product of differential distribution of other factors like socio-economic status of individuals, their dietary and lifestyle habits and access to health care services. In addition, several studies have been conducted on Asian men, specifically those of Chinese and Japanese decent living in Western countries. They show that there are higher incidences of PCa in immigrants compared to their counterparts (Brawley, 2012; Lichtensztajn *et al.*, 2014; Rebbeck, 2017).

Thus, if genetic pre-disposition is the major contributing factor, then the substantial difference in incidence among men of the same origin, living in different countries cannot be explained. These findings thus suggest that there is a possible role of lifestyle and dietary patterns as contributing factors in PCa development. Although genetics play a vital role in PCa development, considerable evidence suggests that environmental factors (mainly diet and lifestyle) are important in the development of PCa (Dall'Era *et al.*, 2018; Li *et al.*, 2018; Nettey *et al.*, 2018). However, gene-environment interactions must play a central role in the development of the disease.

Findings from autopsy studies suggest the prevalence of small prostate cancers in numerous populations regardless of race or ethnic variations in the incidence (Zlotta *et al.*, 2013; Jahn *et al.*, 2016). The question then arises whether the factors resulting in these slow growing tumours are the same as the factors responsible for the development of larger and clinically significant disease. Thus, even though multiple factors including androgen levels in the body, lifestyle factors and diet are considered to be associated with differential tumour biology among different races, the evidence on all these factors remains unclear (Brawley, 2012).

1.3.7 Hormones as risk factors

There is increasing biological evidence that hormones, particularly androgens play a vital role in the development of PCa. As a result, androgens have remained a mainstay of therapy for advanced PCa (Dai *et al.*, 2017). Testosterone and dihydrotestosterone are the two most important androgens in normal prostate development. They bind to the androgen receptors (AR) and lead to the activation of transcriptional programs that are critical for the maintenance of mature prostate physiologic functions (Griffin, 1992; Green *et al.*, 2012). Thus, signalling via the AR axis is thought to be important in facilitating prostate carcinogenesis, although the precise mechanisms driving initiation and progression of prostate cancer are not well elucidated (Green *et al.*, 2012). However, there have been several postulations put forward suggesting that there are alternative pathways of androgen synthesis from the canonical pathway.

During normal prostate physiological function, the canonical biosynthesis of all steroid hormones begins with a 27-carbon cholesterol, which undergoes stepwise modification by a small complement of enzymes first to 21-carbon steroids (progestins) and subsequently to 19-carbon androgens (Figure 1.2). During the conversion of androstenedione (AD) to dihydrotestosterone (DHT), AD first forms first forms testosterone through reduction of its 17-keto moiety mediated by 17β -HSD, before 5α -reduction to DHT by SRD5A (Figure 1.2). In contrast, an alternative pathway has been described, in which AD can bypass testosterone as an obligate precursor, instead undergoing 5α -reduction to an intermediate 5α -androstenedione (5α -dione), followed by 17-keto reduction to DHT (Chang *et al.* 2011). This pathway is appears to be the favoured mechanism in all prostate cancer cell lines as well as sampled castration resistant prostate cancer biopsies from patients (Chang *et al.*, 2011; Dai *et al.*, 2017).

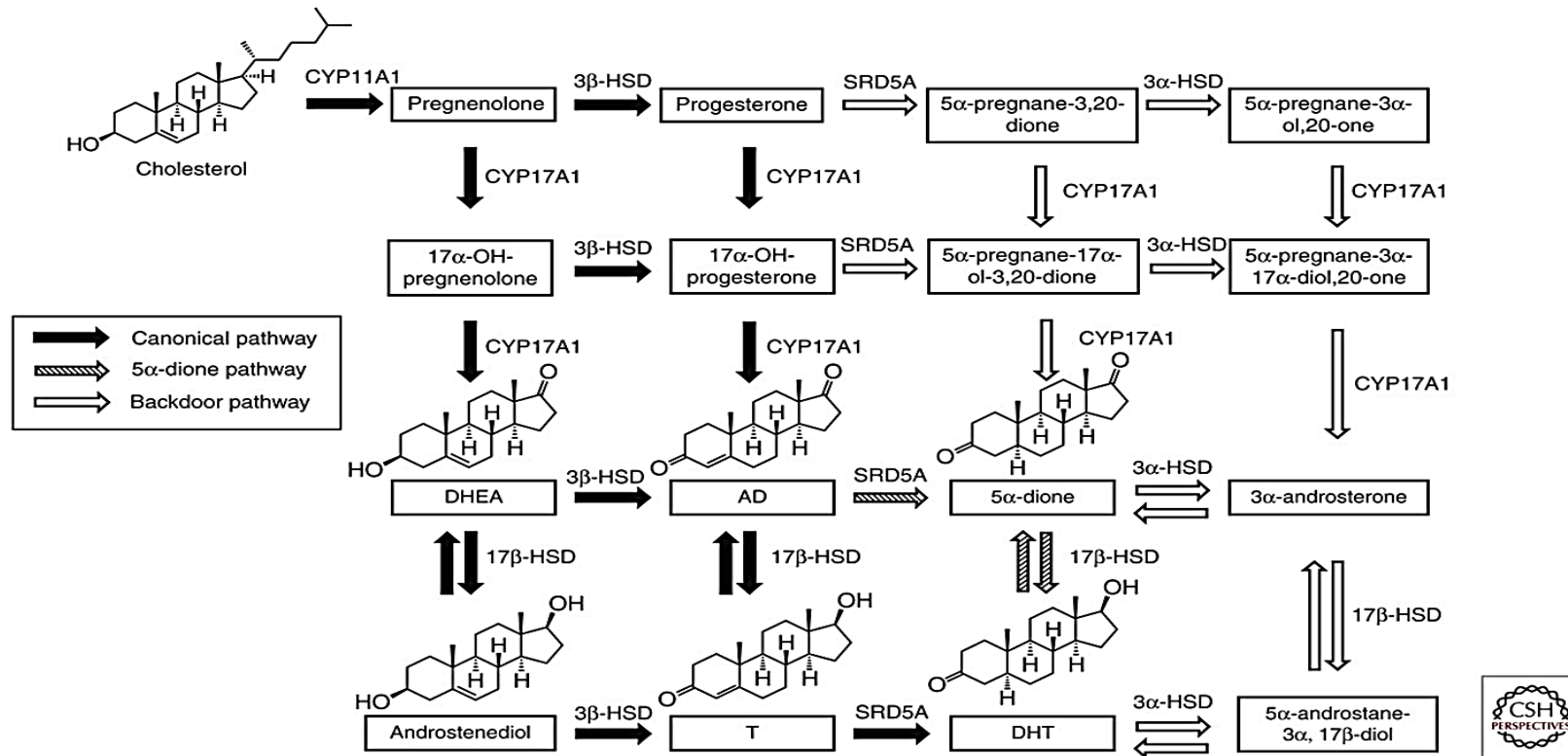


Figure 1.2: Pathways of androgen biosynthesis in normal physiology and prostate cancer. Key enzymes are denoted next to arrows for each reaction. (Adapted from Dai *et al.*, 2017).

1.3.8 Environmental risk factors of prostate cancer

The aetiology of PCa is not well understood. However, the role of environmental factors on the development and progression of PCa has been extensively investigated and reported in several populations. North American populations have the highest incidence rate of the disease with men of African heritage having higher rates than those of European and Asian ancestry respectively (Malik *et al.*, 2018). Some of these statistics are attributed to various environmental influences such as diet, consumption of tobacco and exposure to heavy metals and industrial chemicals (Korc *et al.*, 2017; Darcey and Boyle, 2018; Vineis and Fecht, 2018).

There are currently no reports linking the association of consumption of tobacco smoking and PCa incidence (Darcey and Boyle, 2018). However, there are a number of publications reporting modest associations between tobacco smoking and PCa recurrence, overall mortality and prostate-cancer specific mortality (Islami *et al.*, 2014; Rieken *et al.*, 2015; Ganesh *et al.*, 2016; Darcey and Boyle, 2018). This suggests an impact of tobacco smoking on survival and recurrence of the disease. However, the mechanisms behind these associations are not well documented and poorly understood. Perhaps studies in these aspects could benefit from the use of biomarkers to monitor recurrence and mortality in PCa patients who are tobacco smokers.

Several studies have shown the implication of diet in the development of prostate cancer (Fan *et al.*, 2018; Schneider *et al.*, 2018). Animal fats, dairy and calcium have all been linked to an increased risk of prostate cancer as well as aggressiveness of the disease in several populations including African American and Caucasian populations (Aune *et al.*, 2015; Schwingshackl *et al.*, 2017). Data

in these studies also suggests that a Mediterranean- style diet and a greater consumption of whole grains, vegetables and fruits promotes a protective effect against aggressive PCa.

However, further studies are needed to determine dietary ratios and link to the aggressiveness of the disease. Additionally, these data need to be compared and/or corroborated in African populations, as the socioeconomic features of the continent are different from the American and European aspects which leads to great differences in dietary components.

A number of molecular biomarkers are reported to be involved in prostate tumourigenesis and clinical progression. There is increasing evidence that these markers are targets of essential metals such as zinc, copper and iron (Tan and Chen, 2011; Vella *et al.*, 2017). Additionally, epidemiological studies have shown that exposure to non-essential heavy metals such as cadmium and arsenic may have roles in prostate carcinogenesis (Tan and Chen, 2011). These metals are thought to act via several mechanisms including, ROS generation, deregulation of proliferation, apoptosis and angiogenesis as well as by hormone-mimicking mechanisms (Vella *et al.*, 2017). Heavy metals, such as cadmium (Cd) and arsenic (As) act as endocrine disruptors by interfering with androgen signalling pathways in prostate cancer. The Cd may bind to the androgen receptor and induce proliferation in androgen dependent prostate cells (Vella *et al.*, 2017).

1.4 Screening, diagnosis and prognosis of prostate cancer

Prostate cancer survival is related to a number of factors, including the extent of tumour at the time of diagnosis (Hoffman *et al.* 2018). The five-year relative

survival rate among men with PCa confined to the prostate (localized) is as high as 95 % percent, compared with 29.3 percent among those diagnosed with distant metastases (Ries *et al.*, 2007). While men with advanced stage disease may benefit from palliative treatment, their tumours are generally not curable. Therefore, a screening program that could accurately identify asymptomatic men with aggressive localized tumors might be expected to substantially reduce prostate cancer morbidity, as well as painful metastases, and subsequently mortality.

1.4.1 Prostate specific Antigen

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 high PSA levels are indicative of prostate cancer, as PSA levels are organ specific and not cancer specific (Schröder *et al.*, 2014; Hoffman *et al.*, 2018). 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 the 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 (Hoffman *et al.*, 2018). 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 be less effective with advancing age.

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; Hoffman *et al.*, 2018). Over the last 25 years, no new blood test, genetic test or medical x-ray have been able to replace PSA. However, the 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)

Although digital rectal examinations (DREs) have long been used to diagnose PCa, recent studies suggest not performing the digital rectal examination (DRE) for prostate cancer screening either alone or in combination with prostate-specific antigen (PSA) screening. This is because of a lack of evidence of reduction in the morbidity or mortality of Pca when detected by DRE at any age in controlled studies (Marshall, 2005; Chua *et al.*, 2014; Hoffman *et al.*, 2018).

Additionally, there are a number inherent limitations to the DRE. It can detect palpable abnormalities such as nodules, asymmetry and induration in the posterior and lateral aspects of the prostate gland where the majority of cancers arise (Chua *et al.*, 2014; Kohestani *et al.*, 2018; Hoffman *et al.*, 2018). However, other areas of the prostate where cancer occurs are not reachable by a finger examination (Carter *et al.*, 2013). Furthermore, DREs are dependent on the ability of a doctor to feel the differences between a normal prostate and a tumour. In addition, stage T1 prostate cancers are nonpalpable by definition and are easy to miss (Horwich *et al.*, 2010).

The National Health Institute (NIH) in the UK released 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). Approximately, 54 % of the tested men had induration, marked asymmetry, or nodularity of the prostate. Conclusions from the study indicated that DREs had an estimated sensitivity of 51 %, a specificity of 59 %, and a calculated overall positive predictive value of 41 % for detection of prostate cancer. It was also concluded that DRE screening increased the likelihood of finding early disease. However, the quality of evidence in the study was very low and there was substantial heterogeneity across studies that was not taken into

account. Therefore, the use of DRE screening for prostate cancer remains controversial.

1.4.3 The prostate cancer antigen 3 gene (PCA3)

The PCA3, is highly overexpressed in almost all prostate cancer tissue specimens but not in normal or hypertrophied tissue (Carroll *et al.*, 2001). A PCA3 score, based on the ratio of PCA3 mRNA over prostate-specific antigen (PSA) mRNA (which is not related to serum PSA levels or cancer), can be determined from a urine specimen collected after a vigorous digital rectal examination (Vlaeminck-Guillem *et al.*, 2015; Rodon *et al.*, 2018). Prostate cancer antigen 3 has been evaluated for guiding biopsy decisions when PSA levels are in an indeterminate range (3 to 7.0 ng/mL) and for men with previously negative biopsies but persistently elevated PSA levels (Rodon *et al.*, 2018). Four clinical trials evaluating the diagnostic performance of PCA3 reported a sensitivity range of 53 % to 84 % and a specificity range of 71 % to 80 % (Vlaeminck-Guillem *et al.*, 2010; Bradley *et al.*, 2013). While, another three reported a sensitivity range of 47 % to 58 %, and specificity range of 71 % to 72 % (Vlaeminck-Guillem *et al.*, 2015). However, the latter three studies were performed following a previous negative biopsy (Rodon *et al.*, 2018).

These results indicate that PCA3 may eventually have a role in reducing unnecessary biopsies. However, more data on clinical outcomes are needed to support its routine use (Vlaeminck-Guillem *et al.*, 2015; Rodon *et al.*, 2018). The PCA3 biomarker is currently approved by the US Food and Drug Administration (FDA). However, there is a caveat, its main usage is to inform decisions on

whether to repeat a prostate biopsy in men ≥ 50 years with one or more previous negative biopsies (Hoffman *et al.*, 2018).

1.4.4 Prostate biopsy

A prostate biopsy can be ordered as a follow up on a PSA test, DREs and/or a PCA3 test (Pilatz *et al.*, 2018). 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; Sriplakich *et al.*, 2018). 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 to 18 times (Pilatz *et al.*, 2018). 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 (Pilatz *et al.*, 2018; Sriplakich *et al.*, 2018).

1.4.5 Prostate cancer prognosis

As reviewed in the previous sections, the diagnostics of prostate cancer is complicated due to the heterogeneity and differentiated progression in various subgroups of the disease. These conditions also plague the management of the disease and to date, there is no FDA approved biomarker used to monitor the progression of the disease as well as a therapeutic biomarker (Ali *et al.*, 2018; Chistiakov *et al.*, 2018).

There have been several studies focused around the discovery of and validation of novel prognostic biomarkers to improve the identification of patients at risk of aggressive PCa or of tumour relapse (Ali *et al.*, 2018; Lamy *et al.*, 2018; Chistiakov *et al.*, 2018). There are currently six prognostic markers approved by the FDA for use in PCa cases. These markers are licensed by several co-operations and are available in parts of Europe and the USA (Lamy *et al.*, 2018). Two of the tests are blood based, the Prostate Health Index (PHI), developed by Beckman Coulter (Loeb and Catalona, 2014) and the 4Kscore, developed by OPKO Health (Voigt *et al.*, 2017). One test is urinary based, the MiProstate Score Urine test (MiPS), developed by MLabs (Lebastchi *et al.*, 2017). The remaining three are molecular signature based tests, the Genomic Prostate Score (GPS) by Genomic Health (Klein *et al.*, 2016), Prolaris Cell Cycle Progression score (CCP), by Myriad Genetics (Ontario Health Technology, 2017) and the Decipher Genomic Score (GC), developed by GenomeDx. (Dalela *et al.*, 2016).

Before examining the performance of the biomarkers/tests, it is necessary to understand the ranking system that is used to describe the strength of the results (based on prognosis, diagnosis, economic and decision analysis) measured in a clinical trial or research study. These are level of evidence (LOE), clinical validity and clinical utility (Burns *et al.*, 2011). The level of evidence is a

system of rating evidence (Table 1.1) when determining the effectiveness of a particular intervention (Burns *et al.*, 2011). For the purpose of this research, the prognostic factor is examined.

Table 1.1: Levels of evidence for prognostic studies (Adapted from Burns *et al.*, 2011)

Level	Type of evidence
I	High quality prospective cohort study with adequate power or systematic review of these studies.
II	Lesser quality prospective cohort, retrospective cohort study, untreated controls from an RCT, or systematic review of these studies.
III	Case-control study or systematic review of these studies
IV	Case series.
V	Expert opinion; case report or clinical example; or evidence based on physiology, bench research or “first principles”

The clinical validity corresponds to the biomarker’s discriminatory power and its ability to divide; independently of other classical markers, a group of patients into subgroups based on a given characteristic, such as tumour aggressiveness or clinical recurrence (Bossuyt *et al.*, 2012; Lamy *et al.*, 2018)

The clinical utility of a biomarker corresponds to the additional value relative to the usual markers and to the benefit/risk ratio linked to the use of that biomarker such as the number of unnecessary biopsies avoided and number of aggressive cancers missed (Bossuyt *et al.*, 2012). The clinical utility needs to be significant enough to lead to a change in patient management (Lamy *et al.*, 2018).

Based on the above understanding, an extensive evaluation of the clinical validity, level of evidence and clinical utility of six prognostic markers was done by Lamy *et al.*, 2018. They present evidence that only two of the markers, the blood based PHI and 4Kscore have the ability to predict aggressive prostate cancer and thus could help clinicians in the management of patients with localised PCa. The other biomarkers showed potential prognostic value, however, evaluation in additional studies is required to confirm their clinical validity (Lamy *et al.*, 2018). These markers however are not available on the South American, Asian or African market. Additionally, their application protocols are complicated and include a blood based fluorescence immunoassay coupled with the calculation of an algorithm combining four kallikreins levels and clinical data for the 4Kscore test. Thus, there is a need for an ideal diagnosis or prognosis marker for the African market that is conveniently available to the general population, minimally invasive with a high level of evidence and clinical validity.

1.4.6 The effect of current prognostic biomarkers on management of prostate cancer

1.4.6.1 Active surveillance

The PHI test in conjunction with a tissue biopsy of DNA content is also available as a management option for active surveillance in men with localised PCa (Isharwal *et al.*, 2011; Tosoian *et al.*, 2012). Studies have shown that PHI measured at diagnosis as well as during the surveillance has discriminatory power when compared to the use of PCA3 and TMPRRS2:ERG (Isharwal *et al.*, 2011; Lin *et al.*, 2013; Lamy *et al.*, 2018). The PHI also has a classified LOE of II (Described in Table 1.1). Thus, it improves the accuracy to predict unfavourable biopsies to 70 %. However, the reliance on biopsy procedure for reasons highlighted in Section 1.4.4 can be a deterrent to patients on active surveillance (Isharwal *et al.*, 2011; Hirama *et al.*, 2014). Additionally, there is no data on the clinical utility of PHI as an active surveillance marker (Lamy *et al.*, 2018).

1.4.6.2 Radical prostatectomy

The PHI test has weak discriminatory power for endpoint prediction of tumour aggressiveness based on the Gleason score and staging on prostatectomy specimens (Guazzoni *et al.*, 2012; Fossati *et al.*, 2015). Lamy *et al.*, 2018, corroborated these findings with a LOE assignment of V, which is the lowest in the classification. There are results on PHI as a prognostic marker from other studies which are not concordant with the above findings, with the LOE's of II and III respectively (Ferro *et al.*, 2015; Mearini *et al.*, 2015). However, the

methodologies used were questionable (eg, subjective measurement of tumour volume, non-independent value of the marker, and very low number of events). Regardless of its shortcomings, studies show that the PHI test has additional value compared to classical parameters such as age, digital rectal examination and PSA (Guazzoni *et al.*, 2012; Cantiello *et al.*, 2015). Its influence on decision-making in monitoring patients after radical prostatectomy remains limited in the clinical setting (Lamy *et al.*, 2018).

The 4Kscore test has also been used in conjunction with PSA to screen for metastasis after radical prostatectomy. The test scored a LOE of III as it enhanced prediction of metastasis compared with PSA alone (Stattin *et al.*, 2015). However, the marker had a low clinical validity (Stattin *et al.*, 2015; Lamy *et al.*, 2018).

1.4.6.3 Adjuvant treatment

A study by Ross *et al.*, 2016 evaluated the use of the 4Kscore in a cohort of 545 patients who received adjuvant radiotherapy as a consequence of stage 3 PCa (Ross *et al.*, 2016). They found that the test is unable to predict biochemical recurrence as well as distant metastasis. A later study by Kretschmer *et al.*, 2017 reported similar results (Kretschmer *et al.*, 2017). Ross *et al.* further evaluated the prognostic value of the marker in a cohort of 85 patients with high-risk PCa and biochemical recurrence after radical prostatectomy. They found that 8% of patients with a low-risk profile based on the biomarker developed distant metastases within the follow-up period. They thus concluded that this marker may not be suitable for use as part of the clinical decision-making

process if adjuvant radiotherapy is considered in a high-risk patient (Ross *et al.*, 2016).

It can be seen from the preceding review that the diagnosis, treatment, prognosis and management of PCa continues to be challenging, as the most effective current biomarkers are kallikrein/prostate-specific antigen (PSA) based. The only reliable method for PCa diagnosis and management is the prostate biopsy (Ross *et al.*, 2016). Thus, there is a need to identify a biomarker that could distinguish, malignant cancer from benign prostatic hyperplasia (BPH) during the early diagnosis of the disease as well as localised and metastasised disease.

Studies have shown that aberrantly expressed microRNAs are a hallmark of several diseases such as cancer (Garzon *et al.*, 2014; Hill and Tran, 2018). MicroRNA expression profiling has increasingly been shown to be associated with tumour development, progression and response to therapy in PCa (Fabris *et al.*, 2016; Kanwal *et al.*, 2017), suggesting their possible use as diagnostic, prognostic and predictive biomarkers.

1.5 MicroRNAs

MicroRNAs are naturally occurring endogenous, single stranded RNA molecules. They are 18 -24 nucleotide bases long and are non-protein coding (Macfarlane and Murphy, 2011; Kanwal *et al.*, 2017). They control gene expression by binding to target mRNAs with imperfect complementarity within the 3'-UTR, leading either to repression of translation or degradation (Cannell *et al.*, 2008). MicroRNAs are subject of intensive research and as a result, a wealth of information on their biogenesis, function and significance in gene regulation has

been amassed. MicroRNAs 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 microRNA mutation or expression can result in various disorders including cancer (Macfarlane and Murphy, 2011).

MicroRNAs are known to regulate gene expression at either the transcriptional or the post-transcriptional level (Bartel, 2004). Studies have shown that microRNAs can target enhancers and thereby activate gene expression (Catalanotto *et al.*, 2016). A recent study found that miR-24-1 targets enhancers in the KDM6B gene locus in the nucleus (Xiao *et al.*, 2017). The study also found that this ability does not take away from the capacity of the microRNA to function canonically by repressing its target at the mRNA level in the cytoplasm (Xiao *et al.*, 2017).

1.5.1 MicroRNAs and microRNA target genes as network biomarkers for PCa

A number of studies have been published recently on network biomarkers (Wu *et al.*, 2014; Yang *et al.*, 2018). This is because genes associated with a complex disease such as cancer never function alone, but work together in a complex network. Thus, employing network biomarkers and signatures such as microRNAs and their target genes as well as co-expressed genes in biomarker

discovery would greatly improve the methods currently being used (Wu *et al.*, 2014).

A microRNA miR-301a, was used by Damodaran *et al.*, 2016 as a prognostic marker for to differentiate between PCa and benign prostate hyperplasia. They found that overexpression of the microRNA, activates invasion/migration of PCa cells, a pre-requisite for tumour proliferation (Damodaran *et al.*, 2016). Another microRNA, microRNA-141 and its gene targets have recently been used as diagnostic and prognostic markers for prostate cancer in an Egyptian study (Ali *et al.*, 2018). Their study found that the microRNA was significantly overexpressed in PCa patients with metastasis in comparison with those without.

Preliminary studies have also been conducted on microRNAs and their target genes as regulatory pairs for predicting the response to chemoradiotherapy in rectal cancer using a bioinformatics approach (Peng *et al.*, 2017). In the aforementioned study, candidate microRNAs were identified as associated with the response to chemoradiotherapy in rectal cancer using the PubMed platform. An inference was made that if the biomarker microRNAs can predict the response of rectal cancer to preoperative chemoradiotherapy, the mRNAs they regulate should also participate in chemoradiotherapy response and diverse associated biological pathways. Results from this study indicated that the target genes are indeed critical in the response to ionizing radiation and are therefore good candidates markers for radiation therapy (Peng *et al.*, 2017).

1.6 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 (Yan and Yan, 2017) 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 (Khudoshin and Yuryev, 2018; Singh *et al.*, 2019).

Hence, 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 (Singh *et al.*, 2019). 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 (Peng *et al.*, 2017; Singh *et al.*, 2019). 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 microRNAs and their target genes, or proteins for the discovery of possible novel biomarkers for diseases (Sommer *et al.*, 2010; Peng *et al.*, 2017).

Identifying microRNAs, their target genes, and their respective regulatory functions is 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). Using software algorithms to search for characteristic markers allows the scanning of large sets of data in

minutes. It also allows the elimination of biomarkers that do not adhere to a set of detailed criteria (Peng *et al.*, 2017). The various bioinformatics software and algorithms used in this study will be discussed further in their respective chapters.

1.7. Study rationale

The principal involvement of microRNAs in the aetiology and progression of many common diseases indicates that these molecules are significant markers with potential use as diagnostic, prognostic and therapeutic tools. The discovery that microRNAs are detectable and quantifiable in the circulation of diseased persons adds further validity to their potential as biomarkers of disease and possible distinguishers between benign and malignant tumours.

The microRNA hsa-miR-5698 was discovered to be a diagnostic and prognostic marker in PCa in a Master's study; it is able to distinguish between and prostate adenocarcinoma and bone metastasis (Lombe, 2015). Six translational targets of this microRNA were discovered using *in silico* approaches. This was done because it was inferred that if hsa-miR-5698 is a good candidate biomarker for PCa and can distinguish between metastasis and adenocarcinoma, then the genes it targets should also participate in the same associated biological pathways under certain conditions such as PCa. Thus, instead of using one biomarker such as the microRNA or one target gene, a set of candidate network biomarkers (consisting of the microRNA, and its target genes) can be used to more accurately classify the various stages of PCa for improved diagnostics and prognostics.

1.7.1 Aims and Objectives

Aim

Thus, the purpose of this study was to analyse via expression profiling and characterization, the target genes of hsa-miR-5698 in order to determine their ability to act as putative dynamic network biomarkers for prostate cancer. The study was conducted using a combined *in silico* and molecular approach. The specific objectives are outlined as follows:

Objectives

1. Determine the hsa-miR-5698–mRNA expression associations in prostate cancer and their effects on the disease progression using correlation analysis (Chapter 2).
2. Identify and analyse novel sequence motifs in the promoters of the target genes of hsa-miR-5698 to decipher transcription factor binding and thus transcriptional regulation in prostate cancer disease progression (Chapter 3).
3. Build microRNA, transcription factor and gene regulatory networks to determine role of the dynamic network biomarkers in PCa progression (Chapter 3).
4. Use transcriptomic data and online survival analysis software to assess the prognostic value of the dynamic network biomarkers in PCa (Chapter 4).

5. Expression analysis of the candidate biomarkers in PCa cell lines using qPCR (Chapter 5).

1.8 References

1. Ali, R., El Tabbakh, S., El Delgawy, W., Kotb, A., & Desouky, M. N. (2018). microRNA-141 as diagnostic and prognostic biomarker for prostate cancer in Egyptian population: Pilot study. *African Journal of Urology*, 24(4), 347–352. <https://doi.org/10.1016/J.AFJU.2018.11.006>
2. Alvarez-Cubero, M. J. *et al.* (2017) ‘Somatic Mutations in Prostate Cancer: Closer to Personalized Medicine’, *Molecular Diagnosis & Therapy*, 21 (2), pp. 167–178. doi: 10.1007/s40291-016-0248-6.
3. Ardekani, A.M and Naeini, M.M. (2010). The Role of MicroRNAs in Human Diseases. *Advanced Journal of Medical Biotechnology* 2 (4):161-179.
4. Arora, K., Barbieri, C. E. (2018). Molecular Subtypes of Prostate Cancer. *Current Oncology Reports*, 20(8), 58. <https://doi.org/10.1007/s11912-018-0707-9>.
5. Aune, D. *et al.* (2015) ‘Dairy products, calcium, and prostate cancer risk: a systematic review and meta-analysis of cohort studies’, *The American Journal of Clinical Nutrition*. Oxford University Press, 101(1), pp. 87–117. doi: 10.3945/ajcn.113.067157.
6. 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.
7. Berman H, Henrick K, Nakamura H. (2013). Announcing the worldwide Protein Data Bank. *Nature Structural and Molecular Biology*, 10:990-1203.
8. Berry, R., Schaid, D. J., Smith, J.R., French, A.J., Schroeder, J.J., McDonnell, S.K., Peterson, B.J., Wang, Z.Y., Carpten, J.D., Roberts, S.G., Tester, D.J., Blute, M.L., Trent, J.M., Thibodeau, S.N. (2000). Linkage analyses at the chromosome 1 loci 1q24-25 (HPC1), 1q42.2-43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. *American Journal of Human Genetics*, 66 (2):539–46.

9. Blackwelder, R., & Chessman, A. (2018). Prostate Cancer Screening: Shared Decision-Making for Screening and Treatment. Primary Care: *Clinics in Office Practice*. <https://doi.org/10.1016/J.POP.2018.10.012>
10. Bossuyt, P. M. M., Reitsma, J. B., Linnet, K., & Moons, K. G. M. (2012). Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clinical Chemistry*, 58 (12), 1636–1643. <https://doi.org/10.1373/clinchem.2012.182576>
11. Braunhut, B. L., Punnen, S., & Kryvenko, O. N. (2018). Updates on Grading and Staging of Prostate Cancer. *Surgical Pathology Clinics*, 11(4), 759–774. <https://doi.org/10.1016/J.PATH.2018.07.003>
12. Brawley, O. W. (2012). Trends in Prostate Cancer in the United States. *Journal of the National Cancer Institute Monographs*, (45). <https://doi.org/10.1093/jncimonographs/lgs035>.
13. Breyer, J.P., Avritt, T.G., McReynolds, K.M., Dupont, W.D., Smith, J.R. (2012). Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. *Cancer Epidemiological Biomarkers Preview*, 21 (8):1348–53.
14. Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and reconstructive surgery*, 128 (1), 305–10.
15. Cancer Genome Atlas Research Network, T. C. G. A. R. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, 163 (4), 1011–1025. <https://doi.org/10.1016/j.cell.2015.10.025>.
16. Cancer Research UK (2018), Prostate Cancer Risk Factors, Cancer Research UK.
17. Cannell, I.G., Kong, Y.W., Bushell, M. (2008) How do microRNAs regulate gene expression? *Biochemical Society Transactions* 36:1224–1231.
18. Cantiello, F., Russo, G.I., Ferro, M., *et al.* (2015). Prognostic accuracy of Prostate Health Index and urinary prostate cancer antigen 3 in predicting pathologic features after radical prostatectomy. *Urologic Oncology*, 33, 163.e15–23.

19. Carroll, P., Coley, C., McLeod, D., Schellhammer, P., Sweat, G., Wasson, J., Zietman, A., Thompson, I. (2001). Prostate-specific antigen best practice policy-- part I: early detection and diagnosis of prostate cancer. *Medline*, 44, 51, 91-93.
20. Carter, H.B., Albertsen. P.C., Barry, M.J. (2013). Early detection of prostate cancer: AUA guideline. www.auanet.org/common/pdf/education/clinical-guidance/Prostate-Cancer-Detection.pdf.
21. Chang, K.H., Li, R., Papari-Zareei, M., Watumull, L., Zhao, Y.D., Auchus, R.J., Sharifi, N. (2011). Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *PNAS*, 108: 13728–13733.
22. Chang K-H, Li R, Papari-Zareei M, Watumull L, Zhao YD, Auchus RJ, Sharifi N. 2011. Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *Proc Natl Acad Sci* 108: 13728–13733.
23. Chistiakov, D. A., Myasoedova, V. A., Grechko, A. V., Melnichenko, A. A., & Orekhov, A. N. (2018). New biomarkers for diagnosis and prognosis of localized prostate cancer. *Seminars in Cancer Biology*, 52, 9–16. <https://doi.org/10.1016/J.SEMCANCER.2018.01.012>.
24. Chua, M. E., Lapitan, M. C. M., Morales, M. L., Roque, A. B. M., & Domingo, J. K. (2014). 2013 Annual National Digital Rectal Exam Day: impact on prostate health awareness and disease detection. *Prostate International*, 2(1), 31–36. <https://doi.org/10.12954/PI.13039>
25. Crocetti, E., Giusti, F., Martos, C., Randi, G., Dyba, T., & Bettio, M. (2017). Variability of cancer risk within an area: time to complement the incidence rate. *European Journal of Cancer Prevention: the official journal of the European Cancer Prevention Organisation (ECP)*, 26 (5), pp. 442-446.
26. Dai, C., Heemers, H., & Sharifi, N. (2017). Androgen Signaling in Prostate Cancer. *Cold Spring Harbor Perspectives in Medicine*, 7 (9), a030452. <https://doi.org/10.1101/cshperspect.a030452>.

27. Dalela, D., Löppenber, B., Sood, A., Sammon, J., & Abdollah, F. (2016). Contemporary Role of the Decipher® Test in Prostate Cancer Management: Current Practice and Future Perspectives. *Reviews in Urology*, 18 (1), 1–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/27162506>.
28. Dall’Era, M. A., deVere-White, R., Rodriguez, D., Cress, R. (2018). Changing Incidence of Metastatic Prostate Cancer by Race and Age, 1988–2015. *European Urology Focus*. <https://doi.org/10.1016/J.EUF.2018.04.016>
29. Damodaran, C., Das, T. P., Papu John, A. M. S., Suman, S., Kolluru, V., Morris, T. J., Ankem, M. K. (2016). miR-301a expression: A prognostic marker for prostate cancer. *Urologic Oncology: Seminars and Original Investigations*, 34(8), 336.e13-336.e20. <https://doi.org/10.1016/J.UROLONC.2016.03.009>.
30. Drake, B. F., Lathan, C. S., Okechukwu, C. A., & Bennett, G. G. (2008). Racial Differences in Prostate Cancer Screening by Family History. *Annals of Epidemiology*, 18 (7), 579–583. <https://doi.org/10.1016/J.ANNEPIDEM.2008.02.004>.
31. 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.
32. Engelstein, D. (2017) ‘[MOLECULAR MARKERS FOR PROSTATE CANCER].’ *Harefuah*, 156 (5), pp. 318–321. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28551905> (Accessed: 4 November 2018).
33. Fabris, L., Ceder, Y., Chinnaiyan, A. M., Jenster, G. W., Sorensen, K. D., Tomlins, S., Calin, G. A. (2016). The Potential of MicroRNAs as Prostate Cancer Biomarkers. *European Urology*, 70 (2), 312–322. <https://doi.org/10.1016/J.EURURO.2015.12.054>
34. Fan, L. lai *et al.* (2018) ‘Carbohydrate intake and the risk of prostate cancer’, *Clinica Chimica Acta*. Elsevier, 484, pp. 60–71. doi: 10.1016/J.CCA.2018.05.032.

35. Ferro, M., Lucarelli, G., Bruzzese, D. *et al.* (2015). Improving the prediction of pathologic outcomes in patients undergoing radical prostatectomy: the value of prostate cancer antigen 3 (PCA3), Prostate Health Index (PHI) and sarcosine. *Anticancer Research*, 35:1017–23.
36. FitzGerald, L. M. *et al.* (2017) 'Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin-embedded prostate tumours', *Scientific Reports*. Nature Publishing Group, 7(1), p. 17778. doi: 10.1038/s41598-017-18217-w.
37. Fossati, N., Buffi., N.M, Haese, A. *et al.* (2015). Preoperative prostate-specific antigen isoform p2PSA and its derivatives, %p2PSA and Prostate Health Index, predict pathologic outcomes in patients undergoing radical prostatectomy for prostate cancer: results from a multicentric European prospective study. *European Journal of Urology*, 68:132–8.
38. Garzon, R., Marcucci, G., Croce, C.M. (2014). Targeting MicroRNAs in Cancer: Rationale, Strategies and Challenges. *Drug Discovery*, 9, 775-789.
39. Goh, C.L., Schumacher, F.R., Easton, D., Muir, K., Henderson, B., Kote-Jarai, Z., Eeles, R.A. (2012). Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *Journal of Internal Medicine*, 271, pp. 353–365.
40. Green, S. M., Mostaghel, E. A., Nelson, P. S. (2012). Androgen action and metabolism in prostate cancer. *Molecular and Cellular Endocrinology*, 360(1–2), 3–13. <https://doi.org/10.1016/j.mce.2011.09.046>
41. Griffin, J.E. (1992). Androgen Resistance: The clinical and molecular spectrum. *New England Journal of Medicine*, 326: 611–618.
42. Guazzoni, G., Lazzeri, M., Nava, L. *et al.* (2012). Preoperative prostate-specific antigen isoform p2PSA and its derivatives, %p2PSA and prostate health index, predict pathologic outcomes in patients undergoing radical prostatectomy for prostate cancer. *European Urology*, 61: 455–66.

43. Guedes, L. B., Almutairi, F., Haffner, M. C., Rajoria, G., Liu, Z., Klimek, S., Zoino, R., Yousefi, K., Sharma, R., De Marzo, A. M., Netto, G. J., Isaacs, W. B., Ross, A. E., Schaeffer, E. M., Lotan, T. L. (2017). Analytic, Preanalytic, and Clinical Validation of p53 IHC for Detection of TP53 Missense Mutation in Prostate Cancer. *Clinical Cancer Research: an official Journal of the American Association for Cancer Research*, 23(16), 4693-4703.
44. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/J.CELL.2011.02.013>.
45. Health Quality Ontario, H. Q. (2017). Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. Ontario Health Technology Assessment Series, 17(6), 1–75. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/28572867>.
46. 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.
47. Herbst, M., Joubert, E., (2017). Cancer Association of South Africa (CANSA) Fact Sheet. Retrieved from <https://www.cansa.org.za/files/2017/05/Fact-Sheet-Cancer-May-2017-1.pdf>.
48. Heyns, C. F., Fisher, M., Lecuona, A., van Der Merwe, A. (2011). Prostate cancer among different racial groups in the Western Cape: presenting features and management," *South African Medical Journal*, 101(4):267–270.
49. Hirama H, Sugimoto M, Ito K, Shiraishi T, Kakehi Y. (2014). The impact of baseline proPSA-related indices on the prediction of pathological reclassification at 1 year during active surveillance for low-risk prostate cancer: the Japanese multicenter study cohort. *Journal of Cancer Research Clinical Oncology*, 140:257–63.

50. Hjelmborg, J.B., Scheike, T., Holst, K., Skytthe, A., Penney, K.L., Graff, R.E., Pukkala, E., Christensen, K., Adami, H.O., Holm, N.V., *et al.* (2014). The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiological Biomarkers Preview*, 23 (11):2303–10.
51. Hoffman, R., Elmore, J.G., O’Leary, M.P., Melin, J.A. (2018). Screening for prostate cancer. Available from <https://www.uptodate.com/contents/screening-for-prostate-cancer#H1>. (Accessed on June 10, 2018).
52. 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.
53. International Agency for Research on Cancer (IARC), GLOBOCAN 2018, accessed via Global Cancer Observatory. (Accessed on September 15 2018).
54. Isharwal, S., Makarov, D.V., Sokoll, L.J., *et al.* (2011). ProPSA and diagnostic biopsy tissue DNA content combination improves accuracy to predict need for prostate cancer treatment among men enrolled in an active surveillance program. *Urology*, 77, 763.e1–6.
55. J.M. Lucas, C. Heinlein, T. Kim, S.A. Hernandez, M.S. Malik, L.D. True, C. Morrissey, E. Corey., *et al.* (2014). The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumour microenvironment and promotes prostate cancer metastasis. *Cancer Dis*, pp. 1310-1325.
56. Jahn, J. L., Giovannucci, E. L., Stampfer, M. J. (2015). The high prevalence of undiagnosed prostate cancer at autopsy: implications for epidemiology and treatment of prostate cancer in the Prostate-specific Antigen-era. *International Journal of Cancer*, 137(12), 2795–2802. <https://doi.org/10.1002/ijc.29408>

57. Jamsaspishvili, T. *et al.* (2018) 'Clinical implications of PTEN loss in prostate cancer', *Nature Reviews Urology*. Nature Publishing Group, 15(4), pp. 222–234. doi: 10.1038/nrurol.2018.9.
58. Johng, D., Ewing, C. M. and Isaacs, W. B. (2017) 'Abstract 3524: HOXB13 collaborates with AR to alter the cellular phenotype of prostate epithelial cells through cytokeratin 14 upregulation', *Cancer Research. American Association for Cancer Research*, 77 (13 Supplement), pp. 3524–3524. doi: 10.1158/1538-7445.AM2017-3524.
59. Kanwal, R., Plaga, A. R., Liu, X., Shukla, G. C., & Gupta, S. (2017). MicroRNAs in prostate cancer: Functional role as biomarkers. *Cancer Letters*, 407, 9–20. <https://doi.org/10.1016/J.CANLET.2017.08.011>
60. Kearns, J.T., Lapin, B., Wang, E., Roehl, K.A., Cooper, P., Catalona, W.J., Helfand B.T. (2016). Associations between iCOGS single nucleotide polymorphisms and upgrading in both surgical and active surveillance cohorts of men with prostate cancer. *European Urology*, 69 pp. 223–228.
61. Khudoshin, A., Yuryev, A. (2018). Application of bioinformatics for biomarkers research, drugs and clinical practice. *Genes*, 17, S15. <https://doi.org/10.1016/J.MGENE.2018.05.061>
62. Kim, S.D., Park, R.Y., Kim, Y.R., Kim, I.J., Kang *et al.* (2010). HOXB13 is co-localized with androgen receptor to suppress androgen stimulated prostate-specific antigen expression. *Analytical Cell Biology*, 43, pp. 284–293.
63. Kim, E. H., Andriole, G. L., Crawford, E. D., Sjoberg, D. D., Assel, M., Vickers, A. J., & Lilja, H. (2017). Detection of High Grade Prostate Cancer among PLCO Participants Using a Prespecified 4-Kallikrein Marker Panel. *The Journal of Urology*, 197 (4), 1041–1047. <https://doi.org/10.1016/J.JURO.2016.10.089>.
64. Klein, E. A., Zhang, N., Crager, M., Maddala, T., Febbo, P. G., Thomas, S., ... Lawrence, H. J. (2016). A 17-gene genomic prostate score (GPS) as a predictor of biochemical (BCR) and clinical recurrence (CR) in men with surgically treated

- intermediate- and high-risk prostate cancer (PCa). *Journal of Clinical Oncology*, 34(2_suppl), 104–104. https://doi.org/10.1200/jco.2016.34.2_suppl.104.
65. Ko, C.-J., Huang, C.-C., Lin, H.-Y., Juan, C.-P., Lan, S.-W., Shyu, H.-Y., ... Lee, M.-S. (2015). Androgen-Induced TMPRSS2 Activates Matriptase and Promotes Extracellular Matrix Degradation, Prostate Cancer Cell Invasion, Tumour Growth, and Metastasis. *Cancer Research*, 75 (14), 2949–2960. <https://doi.org/10.1158/0008-5472.CAN-14-3297>.
66. Kohestani, K., Chilov, M., & Carlsson, S. V. (2018). Prostate cancer screening- when to start and how to screen? *Translational Andrology and Urology*, 7(1), 34–45. <https://doi.org/10.21037/tau.2017.12.25>.
67. Kretschmer, A., & Tilki, D. (2017). Biomarkers in prostate cancer – Current clinical utility and future perspectives. *Critical Reviews in Oncology/Hematology*, 120, 180–193. <https://doi.org/10.1016/J.CRITREVONC.2017.11.007>.
68. Kulda, V., Topolcan, A., Kucera, R., Kripnerova, M., Srbecka, K., Hora, M., Pesta, M. (2016). Prognostic Significance of TMPRSS2-ERG Fusion Gene in Prostate Cancer. *Anticancer Research*, 36 (9), 4787–4794. <https://doi.org/10.21873/anticancer.11037>
69. Lamy, P.J., Allory, Y., Gauchez, A.S., Asselain, B., Beuzeboc, P., de Cremoux, P., Rébillard, X. (2018). Prognostic Biomarkers Used for Localised Prostate Cancer Management: A Systematic Review. *European Urology Focus*, 4 (6), 790–803. <https://doi.org/10.1016/J.EUF.2017.02.017>.
70. Lebastchi, A. H., Russell, C. M., Helfand, A. M., Osawa, T., Siddiqui, J., Siddiqui, R., Morgan, T. M. (2017). PD07-03 MICHIGAN PROSTATE SCORE (MIPS): AN ANALYSIS OF A NOVEL URINARY BIOMARKER PANEL FOR THE PREDICTION OF PROSTATE CANCER AND ITS IMPACT ON BIOPSY RATES. *Journal of Urology*, 197(4). <https://doi.org/10.1016/j.juro.2017.02.378>.
71. Lee, S.R., Choi, Y.D., & Cho, N.H. (2018). Association between pathologic factors and ERG expression in prostate cancer: finding pivotal networking. *Journal of*

- Cancer Research and Clinical Oncology*, 144 (9), 1665–1683.
<https://doi.org/10.1007/s00432-018-2685-6>.
72. Li, J., Siegel, D. A., King, J. B. (2018). Stage-specific incidence rates and trends of prostate cancer by age, race, and ethnicity, United States, 2004–2014. *Annals of Epidemiology*, 28 (5), 328–330.
<https://doi.org/10.1016/J.ANNEPIDEM.2018.03.001>.
73. Lichtensztajn, D. Y., Gomez, S. L., Sieh, W., Chung, B. I., Cheng, I., & Brooks, J. D. (2014). Prostate cancer risk profiles of Asian-American men: disentangling the effects of immigration status and race/ethnicity. *The Journal of Urology*, 191 (4), 952–956. <https://doi.org/10.1016/j.juro.2013.10.075>.
74. Lin, D.W, Newcomb, L.F., Brown, E.C., *et al.* (2013). Urinary TMPRSS2:ERG and PCA3 in an active surveillance cohort: results from a baseline analysis in the Canary Prostate Active Surveillance Study. *Clinical Cancer Research*, 19:2442–50.
75. Liss, M. A., Chen, H., Hemal, S., Krane, S., Kane, C. J., Xu, J., & Kader, A. K. (2015). Impact of Family History on Prostate Cancer Mortality in White Men Undergoing Prostate Specific Antigen Based Screening. *The Journal of Urology*, 193 (1), 75–79. <https://doi.org/10.1016/J.JURO.2014.07.085>
76. Loeb, S., Catalona, W. J. (2014). The Prostate Health Index: a new test for the detection of prostate cancer. *Therapeutic advances in urology*, 6(2), 74-7.
77. Lythgoe, C., Dynda, D., Alane, S. (2016). Hereditary Prostate Cancer. *Prostate Cancer*, 141–147. <https://doi.org/10.1016/B978-0-12-800077-9.00017-7>
78. Macfarlane, L. A., & Murphy, P. R. (2010). MicroRNA: Biogenesis, Function and Role in Cancer. *Current Genomics*, 11 (7), 537-61.
79. Marshall, J. B. (2005). How Adequate Is Digital Rectal Exam for Prostate Cancer Screening At Colonoscopy? Can Adequacy Be Improved? *Gastrointestinal Endoscopy*, 61 (5), AB260. [https://doi.org/10.1016/S0016-5107\(05\)01368-4](https://doi.org/10.1016/S0016-5107(05)01368-4)

80. Martinez-Gonzalez, L. J. *et al.* (2018) 'Improving the genetic signature of prostate cancer, the somatic mutations', *Urologic Oncology: Seminars and Original Investigations*, 36 (6), p. 312.e17-312.e23. doi: 10.1016/j.urolonc.2018.03.012.
81. Martinez-Gonzalez, L. J. *et al.* (2018) 'Improving the genetic signature of prostate cancer, the somatic mutations', *Urologic Oncology: Seminars and Original Investigations*, 36 (6), p. 312.e17-312.e23. doi: 10.1016/j.urolonc.2018.03.012.
82. Mearini, L., Nunzi, E., Ferri, C., *et al.* (2015). Use of the Prostate Health Index for the detection of aggressive prostate cancer at radical prostatectomy. *Urology International*, 95:390–9.
83. Mehra, R., Kumar-Sinha, C., Shankar, S., Lonigro, R.J., Jing, X., Philips, N.E., *et al.* (2011). Characterization of bone metastases from rapid autopsies of prostate cancer patients. *Clinical Cancer Research*, 17 (12):3924–32.
84. Migliore, L., Coppedè, F. (2002). Genetic and environmental factors in cancer and neurodegenerative diseases. *Mutation Research/Reviews in Mutation Research*, 512(2–3), pp. 135–153.
[https://doi.org/10.1016/S1383-5742\(02\)00046-7](https://doi.org/10.1016/S1383-5742(02)00046-7).
85. Mikropoulos, C., Goh, C., Leongamornlert, D., Kote-Jarai, Z., Eeles, R. (2014). Translating genetic risk factors for prostate cancer to the clinic: 2013 and beyond. *Future Oncology*, 10, pp. 1679–1694.
86. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., *et al.* (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*. 315(1):68–76.
87. Mulholland, D.J., Kobayashi, N., Ruscetti, M., Zhi, A., Tran, L.M., Huang, J., *et al.* (2012). Pten loss and RAS/MAPK activation cooperate to promote EMT and metastasis initiated from prostate cancer stem/progenitor cells. *Cancer Research*, 72 (7), pp. 1878–89.
88. Nelson, W.G., De Marzo, A.M., and Yegnasubramanian, S. (2009). Epigenetic alterations in human prostate cancers. *Endocrinology*, 150 pp. 3991–4002

89. Nettey, O. S., Walker, A. J., Keeter, M. K., Singal, A., Nugooru, A., Martin, I. K., ... Murphy, A. B. (2018). Self-reported Black race predicts significant prostate cancer independent of clinical setting and clinical and socioeconomic risk factors. *Urologic Oncology: Seminars and Original Investigations*, 36(11), 501.e1-501.e8. <https://doi.org/10.1016/J.UROLONC.2018.06.011>.
90. Netto, G. J., Eich, M. L., Varambally, S. (2017) 'Prostate Cancer: An Update on Molecular Pathology with Clinical Implications', *European Urology Supplements. Elsevier*, 16(12), pp. 253–271. doi: 10.1016/j.eursup.2017.10.001.
91. Netto, G. J., Eich, M.L., Varambally, S. (2017) 'Prostate Cancer: An Update on Molecular Pathology with Clinical Implications', *European Urology Supplements. Elsevier*, 16(12), pp. 253–271. doi: 10.1016/j.eursup.2017.10.001.
92. Netto, G.J. (2015). Molecular updates in prostate cancer. *Surgical Pathology Clinic*, 8, pp. 561–580.
93. Parsa, N. (2012). Environmental factors inducing human cancers. *Iranian Journal of Public Health*, 41 (11), 1-9.
94. Peng, Q., Zhu, J., Shen, P., Yao, W., Lei, Y., Zou, L.,Zhu, Y. (2017). Screening candidate microRNA-mRNA regulatory pairs for predicting the response to chemoradiotherapy in rectal cancer by a bioinformatics approach. *Scientific Reports*, 7(1), 11312. <https://doi.org/10.1038/s41598-017-11840-7>.
95. Peto J (2010) Cancer epidemiology in the last century and the next decade. *Nature* 411, pp. 390–395.
96. Pilatz, A., Veeratterapillay, R., Köves, B., Cai, T., Bartoletti, R., Wagenlehner, F., ... Pradere, B. (2018). Update on Strategies to Reduce Infectious Complications After Prostate Biopsy. *European Urology Focus*. <https://doi.org/10.1016/J.EUF.2018.11.009>
97. Price, C. P., Christenson, R. H., Linnet, K., Moons, K. G. M. (2008). Evaluating new diagnostic technologies: perspectives in the UK and US. *Clinical Chemistry*, 54 (9), 1421–1423. <https://doi.org/10.1373/clinchem.2008.108217>.

98. Rebbeck, T. R. (2017). Prostate Cancer Genetics: Variation by Race, Ethnicity, and Geography. *Seminars in Radiation Oncology*, 27(1), 3–10. <https://doi.org/10.1016/J.SEMRADONC.2016.08.002>.
99. Rieken, M. *et al.* (2015) 'Association of Cigarette Smoking and Smoking Cessation with Biochemical Recurrence of Prostate Cancer in Patients Treated with Radical Prostatectomy', *European Urology*. Elsevier, 68(6), pp. 949–956. doi: 10.1016/j.eururo.2015.05.038.
100. Ries, L.A.G., Melbert, D., Krapcho, M., (2007). SEER Cancer Statistics Review, 1975-2004, National Cancer Institute, Bethesda. Available at: http://seer.cancer.gov/csr/1975_2004/. (Accessed on October 16, 2018)
101. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, 161(5):1215–28
102. Rodon, N., Trias, I., Verdú, M., Calvo, M., Banus, J. M., & Puig, X. (2019). Correlation of mRNA-PCA3 urine levels with the new grading system in prostate cancer. *Revista Española de Patología*, 52 (1), 20–26. <https://doi.org/10.1016/J.PATOL.2018.04.003>.
103. Roehl, K. A., Loeb, S., Antenor, J. A. V., Corbin, N., & Catalona, W. J. (2006). Characteristics of Patients With Familial Versus Sporadic Prostate Cancer. *The Journal of Urology*, 176 (6), 2438–2442. <https://doi.org/10.1016/J.JURO.2006.07.159>.
104. Ross, A.E., D'Amico, A.V., Freedland, S.J. (2016). Which: when and why? Rational use of tissue-based molecular testing in localized prostate cancer. *Prostate Cancer Prostatic Disease*, 19 (1) pp. 1-6.
105. Rudichuk, L., Vogel, K. J., Wang, C., Helfand, B. T., & Selkirk, C. G. (2017). Urologists' Current Practices in Screening and Treating Men With a Family History of Prostate Cancer. *Urology*, 99, 180–185. <https://doi.org/10.1016/J.UROLOGY.2016.07.032>.

106. Samuels, Y., Waldman, T. (2010). Oncogenic mutations of PIK3CA in human cancers. *Current topics in microbiology and immunology*, 347, 21-41.
107. Sanda, M. G., Feng, Z., Howard, D. H., Tomlins, S. A., Sokoll, L. J., Chan, D. W., ... Taneja, S. (2017). Association Between Combined TMPRSS2:ERG and PCA3 RNA Urinary Testing and Detection of Aggressive Prostate Cancer. *JAMA Oncology*, 3 (8), 1085. <https://doi.org/10.1001/jamaoncol.2017.0177>
108. Schneider, L. *et al.* (2018) 'Dietary patterns based on the Mediterranean diet and DASH diet are inversely associated with high aggressive prostate cancer in PCaP', *Annals of Epidemiology*. Elsevier. doi: 10.1016/J.ANNEPIDEM.2018.08.012
109. Schröder, F. H., Hugosson, J., Roobol M. J., Tammela, T. L., Zappa, M., Nelen, V., Kwiatkowski, M. *et al.* (2014). Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet*. 384 (9959):2027.
110. Schwingshackl, L. *et al.* (2017) 'Adherence to Mediterranean Diet and Risk of Cancer: An Updated Systematic Review and Meta-Analysis', *Nutrients*. Multidisciplinary Digital Publishing Institute, 9 (10), p. 1063. doi: 10.3390/nu9101063.
111. Selkirk, C. G., Wang, C., Lapin, B., Helfand, B. T. (2015). Family History of Prostate Cancer in Men Being Followed by Active Surveillance Does Not Increase Risk of Being Diagnosed With High-grade Disease. *Urology*, 85(4), 742–747. <https://doi.org/10.1016/J.UROLOGY.2014.10.060>.
112. Shen, L. W., Mao, H. J., Wu, Y. L., Tanaka, Y., & Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochemistry*, 142, 1–10. <https://doi.org/10.1016/J.BIOCHI.2017.07.016>
113. Singh, O., Chang, N.-W., Dai, H.-J., & Jonnagaddala, J. (2019). Translational Bioinformatics Databases. *Encyclopedia of Bioinformatics and Computational Biology*, 1058–1062. <https://doi.org/10.1016/B978-0-12-809633-8.20303-8>.

114. Sommer, C., Held, M., Fischer, B., Huber, W., & Gerlich, D. W. (2013). Bioimage informatics Cell H5: a format for data exchange in high-content screening, 29 (12), 1580–1582. <https://doi.org/10.1093/bioinformatics/btt175>.
115. Spencer, C.C., Su, Z., Donnelly, P., Marchini, J., (2009). Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics*, 5: e1000477.
116. Sriplakich, S., Lojanapiwat, B., Chongruksut, W., Phuriyaphan, S., Kitirattakarn, P., Jun-Ou, J., & Amantakul, A. (2018). Prospective performance of the Prostate Health Index in prostate cancer detection in the first prostate biopsy of men with a total prostatic specific antigen of 4–10 ng/mL and negative digital rectal examination. *Prostate International*, 6 (4), 136–139. <https://doi.org/10.1016/J.PRNIL.2018.02.002>.
117. Stattin, P., Vickers, A.J., Sjoberg, D.D., *et al.* (2015). Improving the specificity of screening for lethal prostate cancer using prostate-specific antigen and a panel of kallikrein markers: a nested case-control study. *European Urology*, 68:207–13.
118. Tang, K.D., Ling, M.T. (2014). Targeting drug-resistant prostate cancer with dual PI3K/mTOR inhibition. *Current Medical Chemistry*, 21 (26):3048–56.
119. Tapia-Laliena, M.A., Korzeniewski, N., Hohenfellner, M., Duensing, S. (2014). High-risk prostate cancer: a disease of genomic instability. *Urology Oncology*, 32 (8), pp. 1101–7.
120. Tollefson, M.K. (2012). Focus on PSA Screening for Prostate Cancer. *Clinical Update Mayo Clinic*, 28.
121. Tomlins, S. A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B. E., ... Chinnaiyan, A. M. (2008). Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* (New York, N.Y.), 10 (2), 177–188. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18283340>.

122. Tosoian, J.J., Loeb, S., Feng, Z., *et al.* (2012). Association of proPSA with biopsy reclassification during active surveillance for prostate cancer. *Journal of Urology*, 188:1131–6.
123. Urry, H. A., Sartorius, R. J., Aldous, C. (2015). South African journal of surgery. : Suid-Afrikaanse tydskrif vir chirurgie. South African Journal of Surgery (Vol. 53). *Association of Surgeons of South Africa*. Retrieved from http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23612015000300015.
124. Vella, V. *et al.* (2017) ‘Recent views of heavy metals as possible risk factors and potential preventive and therapeutic agents in prostate cancer’, *Molecular and Cellular Endocrinology*, 457, pp. 57–72. doi: 10.1016/j.mce.2016.10.020.
125. Vertosick, E. A., Poon, B. Y., & Vickers, A. J. (2014). Relative Value of Race, Family History and Prostate Specific Antigen as Indications for Early Initiation of Prostate Cancer Screening. *The Journal of Urology*, 192 (3), 724–729. <https://doi.org/10.1016/J.JURO.2014.03.032>.
126. Vlaeminck-Guillem, V., Devonec, M., Champetier, D., Decaussin-Petrucci, M., Paparel, P., Perrin, P., & Ruffion, A. (2015). Urinary PCA3 to predict prostate cancer in a cohort of 1015 patients. *Progrès En Urologie*, 25 (16), e1–e8. <https://doi.org/10.1016/J.PUROL.2015.10.006>.
127. Voigt, J. D., Dong, Y., Linder, V., Zappala, S. (2017). Use of the 4Kscore test to predict the risk of aggressive prostate cancer prior to prostate biopsy: Overall cost savings and improved quality of care to the US healthcare system. *Reviews in Urology*, 19 (1), 1–10. <https://doi.org/10.3909/RIU0753>.
128. Wang, S., Zhang, Q., Xu, D., Pan, Y., Lv, Y., Chen, X., Yang, L. (2018). Characterize the difference between TMPRSS2-ERG and non-TMPRSS2-ERG fusion patients by clinical and biological characteristics in prostate cancer. *Gene*, 679, 186–194. <https://doi.org/10.1016/J.GENE.2018.09.006>.

129. Wang, Y., Dai, B., Ye, D. (2015). CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis. *International journal of clinical and experimental medicine*, 8(9), 15708-15.
130. Wang, Z., Wang, Y., Zhang, J., Hu, Q., Zhi, F., Zhang, S., Liang, H. (2017). Significance of the TMPRSS2:ERG gene fusion in prostate cancer. *Molecular Medicine Reports*, 16 (4), 5450–5458. <https://doi.org/10.3892/mmr.2017.7281>.
131. Weinberg, R.A. (2013). *The Biology of Cancer*. 2 nd. Edition. Garland Science.
132. Wu, X., Chen, L., & Wang, X. (2014). Network biomarkers, interaction networks and dynamical network biomarkers in respiratory diseases. *Clinical and Translational Medicine*, 3, 16. <https://doi.org/10.1186/2001-1326-3-16>.
133. Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., Yu, W. (2017). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biology*, 14(10), 1326. <https://doi.org/10.1080/15476286.2015.1112487>.
134. Xu J, Lange EM, Lu L, Zheng SL, Wang Z. *et al.*, (2013). HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genetics*, 132(1):5–14.
135. Yan, Q., Yan, Q. (2017). Applying Translational Bioinformatics for Biomarker Discovery. *Translational Bioinformatics and Systems Biology Methods for Personalized Medicine*, 55–65. <https://doi.org/10.1016/B978-0-12-804328-8.00005-X>.
136. Yang, B., Li, M., Tang, W., Liu, W., Zhang, S., Chen, L., & Xia, J. (2018). Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nature Communications*, 9(1), 678. <https://doi.org/10.1038/s41467-018-03024-2>.
137. Zheng, S. L., Evanston, I.L., Rong, N. (2017). CHEK2 mutations increase risk for prostate cancer but do not differentiate risk of lethal from indolent disease. *Journal of Urology*, Vol. 197, No. 4S.

138. Zlotta, A. R., Egawa, S., Pushkar, D., Govorov, A., Kimura, T., Kido, M., van der Kwast, T. H. (2013). Prevalence of Prostate Cancer on Autopsy: Cross-Sectional Study on Unscreened Caucasian and Asian Men. *JNCI: Journal of the National Cancer Institute*, 105 (14), 1050–1058. <https://doi.org/10.1093/jnci/djt151>.

Chapter 2

MicroRNA and Target Gene Correlation analysis

2. 1 Introduction

There has been compelling evidence that microRNA expression is dysregulated in human cancers. This is accomplished through various mechanisms including the amplification or deletion of microRNA genes (Hayashita *et al.*, 2006; Tagawa and Seto, 2005), abnormal transcriptional control of microRNAs (Wang *et al.*, 2014), dysregulated epigenetic changes (Lehmann *et al.*, 2008; Donzelli *et al.*, 2015), and defects in the microRNA biogenesis machinery (Walz *et al.*, 2015). Thus, dysregulated microRNAs have been shown to affect the hallmarks of cancer by, sustaining proliferative signalling, evading growth suppressors, resisting cell death, activating invasion and metastasis, and inducing angiogenesis (Hanahan and Weinberg, 2011). Depending on their target genes, microRNAs can function as either oncogenes or tumour suppressor under certain conditions.

MicroRNAs were previously thought to regulate gene expression in a solely negative manner, including translational inhibition, mRNA sequestration and mRNA de-adenylation and degradation (Bartel, 2004; Bartel, 2009). However, there is increasing evidence indicating that microRNAs oscillate between repression and stimulation in response to specific cellular conditions, sequences and cofactors (Vasudevan, 2012). Additionally, alternative mechanisms of microRNA target recognition at the promoter level have also been elucidated for transcriptional gene silencing (Majid *et al.*, 2010; Catalanotto *et al.*, 2016). These

findings thus indicate there is still a lot to understand about how microRNAs regulate gene expression.

2.1.2 MicroRNAs form triplexes with double stranded DNA at sequence specific binding sites to alter gene expression

MicroRNAs may function to regulate gene expression by binding to the promoter regions of certain genes (Li, 2006; Janowski, 2007; Check, 2007). This is done via the binding of the microRNA to the major groove of the DNA duplex forming structures called triplexes. The binding is via a non-Watson-Crick base pairing termed a Hoogsteen bond (Hoogsteen, 1963; Nikolova, 2011). A Hoogsteen hydrogen bond forms between the N7 position of the purine base, which acts as the hydrogen bond acceptor and the C6 amino group (Figure 2.1 A) of the pyrimidine base which is the hydrogen bond donor, which is subsequently bound to another purine via Watson-Crick pairing in the N3–C4 positions (Hoogsteen, 1963; Nikolova, 2013).

Triplexes can also be formed by reverse Hoogsteen bonds. In this instance, the base not involved in the formation of the double helix is rotated 180° with respect to the one involved in the binding. Additionally, two types of triplexes can be formed based on the orientation of the third strand (5'- 3' or 3'-5'). Parallel triplex structures are formed by TA•U and CG•C⁺ triplets and antiparallel triplexes are formed by TA•A, and CG•G. In the case of CG•C⁺, an acidic condition is favourable for triplex formation because the protonation of cytosine is required for effective triplet formation (Li *et al.*, 2016).

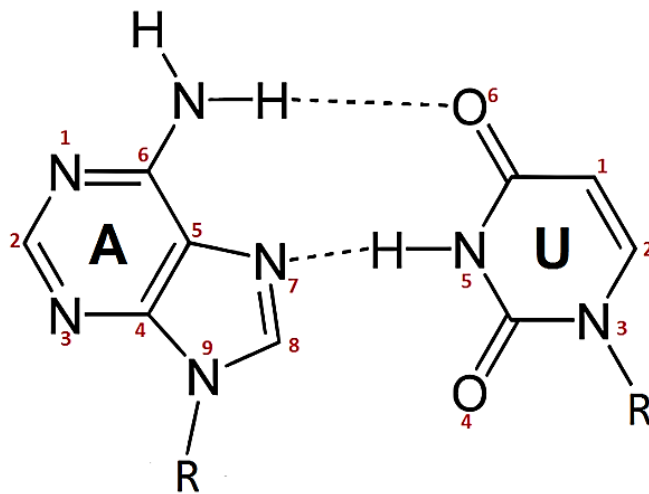


Figure 2.1: Hoogsteen base pair between adenine and uracil (A).

The Hoogsteen hydrogen bonding by the microRNA is weaker than the Watson-Crick hydrogen bonding in the duplex structure. Therefore, there are some limitations to the formation of the triplex structure, especially under physiological conditions such as the acidic conditions mentioned above (Duca *et al.*, 2008).

Given that triplex formation requires base-pairing that is weaker than Watson-Crick base pairing or unusual protonation states for the bases involved, it is significant to consider the existence of such interactions *in vivo*. There are a number of studies making use of bioinformatics tools that suggest the potential for a large presence of DNA-microRNA triplexes in the human genome (Buske *et al.*, 2012; Paugh, 2016).

Interestingly, these motifs tend to accumulate in the gene-regulatory regions, particularly in promoter regions, leading to a proposal that triplex formation might play significant regulatory roles *in vivo*. More direct evidence of triplex formation *in vivo* comes from using triplex-binding antibodies, as well as

fluorescent molecules that selectively recognize triplex structures (Lubitz *et al.*, 2010; Gorab and Pearson, 2018).

This chapter will use computational methods to determine the hsa-miR-5698–mRNA expression associations in prostate cancer and their effects on the disease using correlation analysis. Further, the chapter will investigate whether hsa-miR-5698, has binding sites in the promoters of its translational target genes to form hetero-triplex structures, thereby implicating it in transcriptional control. In chapter three, the thesis will go to use the information obtained in this chapter to build regulatory networks in prostate cancer that may help to understand cell and gene regulation during the initiation and progression during the disease.

2.1.3 Pearson Correlation Analysis

A correlation analysis is used to quantify the association between two continuous variables. That is, between an independent and a dependent variable or between two independent variables. In this case, the former holds, as the microRNA is taken to be the independent variable because it targets the 3' UTR of the mRNA of the target genes preventing gene expression into protein, and consequently, the gene (mRNA) is the dependent variable as its expression may be dependent on the action of the microRNA. Understanding the relationships between the early stages of prostate cancer, its development, metastasis and the action of cellular microRNAs, is crucial in the development of therapeutic strategies in the management of the disease.

There have been several studies (both experimental and computational) conducted that examine the correlation between the expression profiles of

microRNAs and their target genes under various conditions (Huang *et al.*, 2007; Wang and Li, 2009; Kang *et al.*, 2017). In both the experimental and computational studies, results indicate that mRNAs targeted for degradation by microRNAs exhibit an inverse relationship (negative correlation) (Baskerville and Bartel, 2005; Wang *et al.*, 2007; Diaz *et al.*, 2015). On the other hand, intronic microRNAs show a positive expression with their host genes (Wang *et al.*, 2007; Diaz *et al.*, 2015). In this study, a Pearson Correlation analysis will be performed using the formula 2.1 given below.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (2.1)$$

Where,

- X is the expression profile of hsa-miR-5698 in each cell line LNCaP and PC3 and;
- Y is the expression profile of each gene respectively in each cell line LNCaP

In a correlation analysis, a sample correlation coefficient is estimated, which is denoted as *r*. This coefficient ranges between -1 and +1 and quantifies the direction and strength of an association between two variables (microRNA and mRNA expression profiles in PCa) (Benesty *et al.*, 2009).

The sign of the correlation coefficient indicates the direction of the association between the two variables, whilst the magnitude of the correlation coefficient indicates the strength of the association (Benesty *et al.*, 2009; Wang, 2013). For example, a correlation of *r* = 0.9 suggests a strong, positive association between

two variables, whereas a correlation of $r = -0.2$ suggest a weak, negative association. A correlation close to zero suggests no linear association between the variables (Wang, 2013).

2.2 Aims and objectives

2.2.1 Previous work

The microRNA, hsa-miR-5698 was found to be up-regulated in prostate cancer cell lines LNCaP and PC3 in a previous Master's study (Lombe, 2015). A set of bioinformatics tools were used to predict the 3' UTR targets of this microRNA in order to best understand its role in prostate cancer onset and progression and the predicted targets were as follows, CDKN1A/p21, CTNND1, ELK1 BIRC2, FOXC1 and LRP8 (Lombe, 2015).

The chapter aims to investigate if any, the role of hsa-miR-5698 as a transcriptional regulator of the promoters of its translational targets. This is because previous studies (Buske *et al.*, 2011; Paugh, 2016) have shown that binding of microRNAs to gene promoters may result in gene activation by up-regulation of mRNA transcription levels. Thus, the aim is to investigate the dual role of hsa-miR-5698 in the regulation of gene expression in PCa and the implications of this in disease progression and metastasis.

Consequently, this chapter also aims to investigate the distribution of positive and negative correlations between the expression of hsa-miR-5698 and the expression of its translational targets in the prostate adenocarcinoma cell line

LNCaP and the metastatic cell line PC3. This is because genes containing sequences favouring microRNA triplex formation are markedly enriched (3.3 fold, $p < 2.2 \times 10^{-16}$). That is their expression is positively correlated with expression of microRNAs targeting the triplex binding sequences.

Specific objectives:

1. Identify any putative transcriptional effects of hsa-miR-5698 on its translational targets' promoter sequences using the Trident algorithm and PyMOL.
2. Determine whether there is enriched triplex formation for genes whose expression is positively correlated with the expression of hsa-miR-5698 and the implications in PCa progression.
3. Calculate the expression relationship between hsa-miR-5698 and its targets in LNCaP and PC3 cell lines using Pearson Correlation analysis in R Studio.
4. Use the obtained information as well as network inference to build hsa-miR-5698-mRNA putative regulatory networks responsible for primary to metastasis transitions in prostate cancer (Chapter 3).

2.3 Methodology

2.3.1 Identification of putative transcriptional effects of hsa-miR-5698 on its translation targets' promoter sequences.

2.3.1.1 Extraction of gene promoters from the Ensembl database

The promoters sequences were extracted for the set of 6 target genes of hsa-miR-5698 using the Ensembl Genome Browser 91 and saved in FASTA format. The promoter sequences 1000 bp upstream and 200 bp downstream from the transcription start site were extracted in Ensembl by pasting the official gene ID individually into the search box provided on the website and selecting the return button. The configure settings in Ensembl were used to adjust the flanking positions around the transcription start site (TSS). Each of the retrieved promoters was 1000 bp long.

To make certain that the promoter sequences obtained from Ensembl were bonafide, the BLAT tool of the UCSC Genome Browser (accessed at <http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) was used to align them to the sequences annotated in the UCSC database and checking for matches. The BLAT tool also has a CpG islands feature that was turned on to further verify that the promoter sequences were true.

2.3.1.2 MicroRNA triplex binding sites on promoters of target genes

To assess the landscape of potential microRNA triplex binding sites on the promoters of the target genes, the database Trident accessed at <http://trident.stjude.org> was used. This database makes use of a computational algorithm that identifies Hoogsteen and

reverse Hoogsteen interactions between single stranded oligonucleotides such (microRNAs) and double-stranded oligonucleotides (duplex promoter DNA). The results obtained are tabulated in table 2.2.

2.3.2 Visualization of microRNA and promoter DNA triplex

The predicted microRNA-DNA triplex formation obtained from Trident was visualised using PyMOL version 2.2 available at <https://pymol.org/2/>.

2.3.3 Correlation of expression profiles between hsa-miR-5698 its mRNA targets in LNCaP and PC3 cell lines.

The data used in this study was generated in a previous masters study (Lombe, 2015). The expression data for the mature microRNA, hsa-miR-5698 in the LNCaP and PC3 cell lines were generated via qPCR (Lombe, 2015) and the mRNA expression data was generated as described in section 5.2.1. Both sets of data were normalised as described by Pfaffl (2004). Statistical analysis was performed using the R Studio software (RStudio Team, 2015). The Pearson Correlation Coefficients and p-values were computed and used to test the association between the expression profiles of hsa-miR-5698 and its predicted targets in LNCaP and PC3 cell lines.

2.4 Results and discussion

2.4.1 hsa-miR-5698 as a possible transcriptional target of its translational target genes

The landscape of the potential hsa-miR-5698 triplex binding sites on the promoters of its target genes was assessed using the Trident algorithm to identify Hoogsteen and reverse Hoogsteen interactions. For each detected triplex binding site, a thermodynamic binding energy and heuristic score was determined (Table 2.1), with higher heuristic score and a lower thermodynamic energy indicating stronger interaction. The heuristic score is determined based on the number of triplex forming pairs found between hsa-miR-5698 and the helix.

The Trident algorithm was able to predict interaction between hsa-miR-5698 and two genes the set of six genes, namely CDKN1A (p21) and FOXC1. The gene FOXC1 gave a higher heuristic score (164.0) and lower thermodynamic score of -342.70 kJ indicating a stronger interaction between the single oligonucleotide and the DNA. The number of bases participating in the pairing between the two was 14 (Figure 2.3 and Table 2.1).

The p21 gene had a lower heuristic score of 148.0, the thermodynamic score was significantly higher at -198.28 kJ, eight bases on the promoter participated in the pairing. In both cases, parallel triplex structures are formed by Hoogsteen base pairing. The reverse sequence of hsa-miR-5698, an antisense strand (negative control) could not maintain favourable reverse Hoogsteen base pairing, with only four interactions between the microRNA and the helix when compared to the eight formed by the sense strand for the p21 promoter region (Figure 2.2). In

comparison, the antisense strand could not maintain any Hoogsteen base pairing with the FOXC1 promoter region. The results thus indicate that there is enriched triplex formation between the FOXC1 promoter and hsa-miR-5698. This result is interesting because this gene has been shown to be involved in increased tumour size, stromal invasion, and lymph node metastasis in androgen-independent PCa (van Der Heul-Nieuwenhuijsen et al., 2009; Han *et al.*, 2017). Additionally, there is positive correlation between the expression profile of this gene and hsa-miR-5698 in PC3 (Figure 2.4) which is an androgen independent metastasis cell line of prostate cancer.

Table 2.1: Putative interaction of hsa-miR-5698 and the promoters of its translational targets. (Generated with Triden

Gene	hsa-miR-5698 promoter Interaction	Type of match	Disease Type	Binding Strand	Position on promoter	Gibbs Free energy (kJ)	Heuristic Score	Hit structure
CDKN1A/p21	Yes	Direct	Prostate	Sense	853-874	-198.28	148.0	<pre> GGUGUUAGUGACGUGAGGGGGU CACCGCACTCTGGGGAGGGGGC GTGGCGTGAGACCCCTCCCCCG </pre>
CTNND1	No	N/A	Prostate	N/A	N/A		N/A	
BIRC2	No	Direct	Prostate	Sense	N/A		N/A	
FOXC1	Yes	Direct	Prostate	Sense	453-476	-342.70	164.0	<pre> GGUGUUAGUGACGUGAGG--GGGU GGCGGCGGGGAGGAGAGGCTGGGG CCGCCGCCCTCCTCTCCGACCC </pre>
ELK1	No	N/A	Prostate	N/A	N/A		N/A	N/A
LRP8	No	N/A	Prostate	N/A	N/A		N/A	N/A

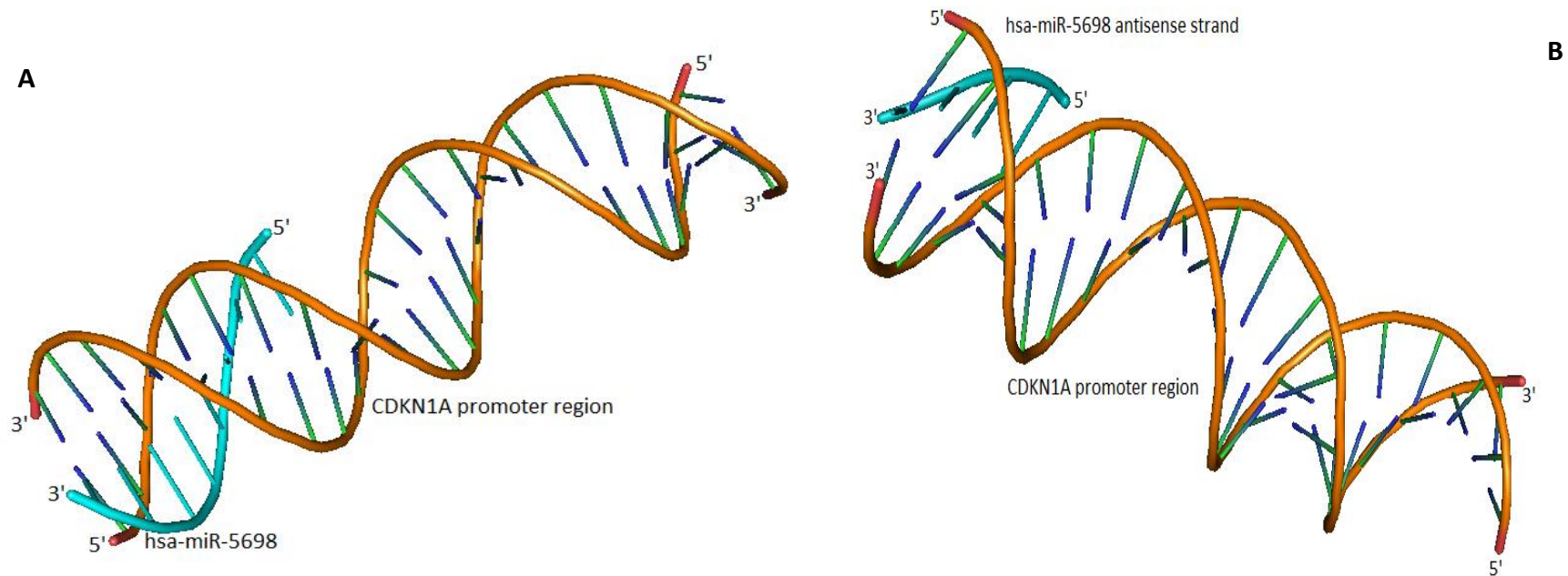


Figure 2.2: Molecular model of hsa-miR-5698-DNA triplex. **(A):** the model predicted for the interaction of hsa-miR-5698 (cyan strand) and the promoter of the gene CDKN1A (p21) showing 8 favourable Hoogsteen base pairings. **(B):** negative control (antisense hsa-miR-5698) (cyan strand) showing less interactions; 4 favourable Hoogsteen base pairings. Both microRNA and DNA duplex are largely twisted and nearly all predicted Hoogsteen pairings cannot be stably maintained. The microRNA bases not involved in the pairing are not shown.

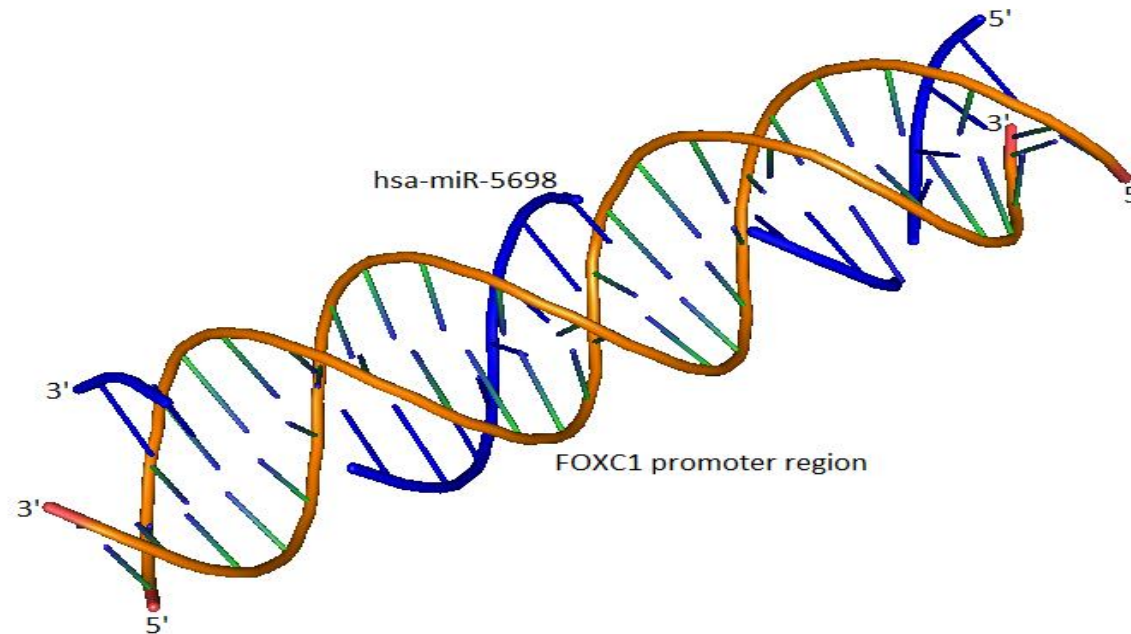


Figure 2.3: Molecular model of hsa-miR-5698-DNA triplex. The model predicted for the interaction of hsa-miR-5698 and the promoter of the gene FOXC1 showing 14 favourable Hoogsteen base pairings. Both microRNA and DNA duplex are largely twisted and nearly all predicted Hoogsteen pairings cannot be stably maintained. The microRNA bases not involved in the pairing are not shown. The antisense of hsa-miR-5698 could not yield a structure for the negative control.

2.4.2 Correlation between hsa-miR-5698 its targets expression profiles in LNCaP and PC3 cell lines.

The correlation analysis at the $P < 0.05$ level in the PC3 cell line revealed a positive relationship (0.84) between FOXC1 and hsa-miR-5698 (Table 2.2). The table also shows that all the other gene expression profiles were negatively correlated to the hsa-miR-5698 expression profile with CDKN1A, CTNND1 and ELK1 showing a moderate negative linear relationship and BIRC2 and LRP8 showing a weak negative relationship. Table 2.2 and Figure 2.4 also show a negative correlation between the expression profiles of hsa-miR-5698 and all the gene expression profiles in the LNCaP cell line. The gene p21 expression profile showed a strong negative linear relationship to hsa-miR-5698, whilst CTNND1 and LRP8 had a moderate negative relationship and BIRC2, FOXC1 and ELK1 showed a weak negative correlation. All correlations conducted in this study were significant (Table 2.2).

Table 2.2: hsa-miR-5698 and their corresponding genes show correlating expression profiles.

Gene	Gene/hsa-miR-5698 in PC3	p-value	Gene/hsa-miR-5698 in LNCaP	p-value
CDKN1A	-0,60	4.27979E-18	-0,95	0.000168359
CTNND1	-0,63	0.001341139	-0,66	7.36352E-14
BIRC2	-0,24	2.08821E-16	-0,17	9.44111E-09
FOXC1	0,84	1.54113E-22	-0,25	2.02307E-16
LRP8	-0,12	4.16011E-13	-0,53	3.02294E-22
ELK1	-0,57	4.18344E-12	-0,14	1.99265E-13

Negative correlation between the expression profiles of the hsa-miR-5698 and the genes accounts for up to 92 % of the analysis conducted. This could indicate that negative regulation is dominant in hsa-miR-5698-mRNA relationship in PC3 and LNCaP. However, an *in vivo* study would have to be conducted to validate these results. There have been a number of studies that have conducted similar analyses to explore microRNA–mRNA expression relationships, but their results were inconsistent. In a correlation analysis on a different set of microRNA and target genes in human bone cells, Laxman *et al.*, found more positive rather than negative correlations (Laxman *et al.*, 2015).

Correlation analysis of hsa-miR-5698 and targets expression profiles

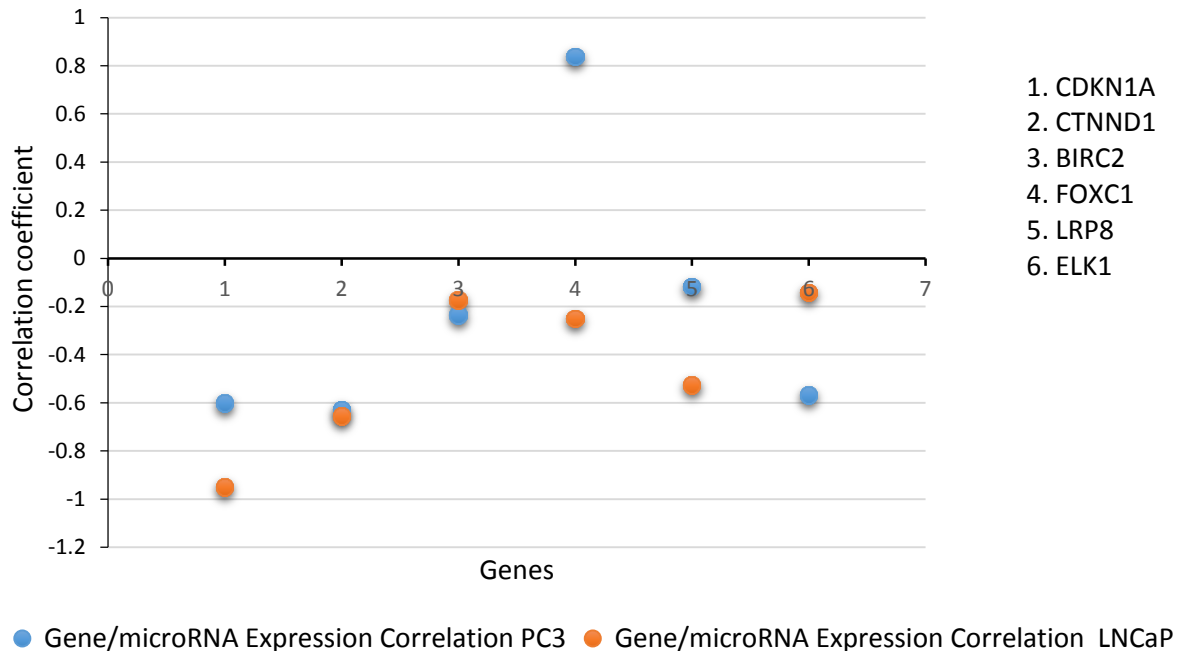


Figure 2.4: Correlation analysis showing the strength of the relationship between hsa-miR-5698 and its targets in the PC3 and LNCaP prostate cancer cell lines.

In another study, Wang and Li (2009) highlighted the effect of different microRNA target prediction tools on correlation analysis. MicroRNA-mRNA target pairs predicted with miRBase showed 46.2 % negative correlation and those predicted by TargetScanHuman appeared 14 % higher at 60.7 % (Wang and Li 2009). The targets in this study were predicted by both tools (Lombe, 2015). A number of positive correlations have been detected by Vasudevan *et al.*, 2007, Nunez-Iglesias *et al.*, 2010 and most recently Wang *et al.*, 2018.

It was further interestingly found that the genes whose expression profiles were found to be negatively correlated with the expression profile of hsa-miR-5698 are enriched in the gene ontology, biological process (BP) related to gene expression regulation (Lombe, 2015). The finding suggests that hsa-miR-5698 could interact with gene expression regulators to indirectly regulate downstream gene expression.

2.5 Conclusion

This study used the Trident algorithm to search for potential hsa-miR-5698 triplex-forming sites in the promoters of the microRNA's translational targets. This was done to understand the dual action of the microRNA in gene regulation in PCa progression. The software PyMOL was used to visualize the binding. It was found that the promoter regions of the genes CDKN1A (p21) and FOXC1 had binding sites for hsa-miR-5698, with the former sustaining eight Hoogsteen base pairs and the latter 14 Hoogsteen base pairs. The study investigated the expression relationship between hsa-miR-5698 and its targets in LNCaP and PC3 cell lines using Pearson Correlation analysis.

Thus, results indicated hsa-miR-5698 regulation of mRNA expression for CDKN1A and FOXC1 might occur through a cis-regulatory effect. That is, the microRNA may bind to the promoter of the gene and result in transcriptional activation of the gene. The microRNA may also have a trans-regulatory effect on the two genes as well as on the other four genes in the study set (no triplex formation). This is in light of the fact that it is well

documented that microRNA regulation occurs mainly at the post-transcriptional mRNA level, through binding to 3'UTR region of complementary mRNAs rather than at the transcriptional DNA level. The potential limitation to this study is that the hsa-miR-5698–mRNA pairs used were predicted and have yet to be validated by cellular and molecular experiments.

It was also observed that there was enriched triplex formation between the microRNA and FOXC1 (Figure 2.3 and Table 2.1). This was consistent with previously recorded findings that positively correlated gene-microRNA expression profiles form more Hoogsteen base pair bonds during triplex formation. From Section 5.3, it was seen that FOXC1 is down-regulated in LNCaP, but up-regulated in PC3 (Figure 5.1). Thus up-regulation of this gene may be cardinal to the progression of PCA to metastasis.

In conclusion, this study examined the correlations between hsa-miR-5698 expression and six of its predicted translational mRNA targets, as well as the potential for the microRNA to regulate these targets via a transcriptional effect. The significant negative correlations discovered and the formation of the triplex between the microRNA and CDKN1A and FOXC1, are evidence that there are cis-regulatory interactions between hsa-miR-5698 and these genes in PC3 and LNCaP. However, the study should be improved by performing luciferase assays to determine the true targets of the microRNA. An electrophoresis mobility shift assay (EMSA) and modified chromatin immunoprecipitation assay (ChIP) can be employed to validate the physical interactions between hsa-miR-5698 and the promoter regions.

The relationship between the gene promoters and hsa-miR-5698 in prostate cancer regulatory networks may serve as a future research target

for an in-depth investigation of the microRNA's biological functions in the disease as well as for better understanding of its molecular pathogenesis in the cancer.

2.6 References

1. Bartel, D. P. (2009). MicroRNA Target Recognition and Regulatory Functions. *Cell*, 136(2), 215–233. <http://doi.org/10.1016/j.cell.2009.01.002>.
2. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281–297.
3. Baskerville, S., Bartel, D.P. (2005). Microarray profiling of microRNAs reveals frequent co-expression with neighbouring microRNAs and host genes. *RNA*, 11 (3), 241-247.
4. Benesty, J., Chen, J., Huang Y., Cohen, I. (2009). Pearson Correlation Coefficient. In: Noise Reduction in Speech Processing. *Springer Topics in Signal Processing*, 2. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00296-0_5.
5. Buske, F.A., Mattick, J.S., Bailey, T.L. (2011) Potential in vivo roles of nucleic acid triple-helices. *RNA Biol* 8: 427–439. pmid:21525785.
6. Catalanotto, C., Cogoni, C., & Zardo, G. (2016). MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *International Journal of Molecular Sciences*, 17 (10), 1712. <http://doi.org/10.3390/ijms17101712>.
7. Chang, T.C., Yu, D., Lee, Y.S., Wentzel, E.A., Arking, D.E., West, K.M. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. *Nature Genetics*, 40, 43–50.
8. Check, E. (2007) RNA interference: hitting the on switch. *Nature* 448: 855–858. PMID 17713502.
9. Diaz, G., Zamboni, F., Tice, A., Farci, P. (2015). Integrated ordination of microRNA and mRNA expression profiles. *BMC Genomics*, 16767 <https://doi.org/10.1186/s12864-015-1971-9>.

10. Donzelli S, Mori F, Bellissimo T, Sacconi A, Casini B, Frixia T et al. (2015). Epigenetic silencing of miR-145-5p contributes to brain metastasis. *Oncotarget*, 6: 35183–35201.
11. Duca, M., Vekhoff, P., Oussedik, K., Halby, L., & Arimondo, P. B. (2008). The triple helix: 50 years later, the outcome. *Nucleic Acids Research*, 36(16), 5123-38.
12. Nikolova, E.N.K., Wise, A.A., O'Brien, P.J., Andricioaei, I., Al-Hashimi, H.A. (2011). "Transient Hoogsteen base pairs in canonical duplex DNA". *Nature*. 470: 498–502.
13. Buske, F.A., Bauer, D.C., Mattick, J.S., Bailey, T.L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Research*, 22, pp. 1372-1381
14. Buske, F.A., Mattick, J.S., Bailey T.L. (2011). Potential in vivo roles of nucleic acid triple-helices. *RNA Biology*, 8, pp. 427-439
15. Gorab, E., Pearson, P.L. (2018). Thiazole Orange as an Alternative to Antibody Binding for Detecting Triple-helical DNA in Heterochromatin of *Drosophila* and *Rhynchosciara*. *Journal of Histochemistry Cytochemistry*, 66 (3), 143-154.
16. Han, L., Witmer, P.D., Casey, E., Valle, D., Sukumar, S. (2007). DNA methylation regulates MicroRNA expression. *Cancer Biology Therapeutics*, 6, 1284–1288.
17. Han, B., Bhowmick, N., Qu, Y., Chung, S., Giuliano, A. E., & Cui, X. (2017). FOXC1: an emerging marker and therapeutic target for cancer. *Oncogene*, 36 (28), 3957-3963.
18. Hanahan, D., Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144: 646–674.
19. Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., et al. (2005). A polycistronic microRNA cluster, miR-17-92, is

- overexpressed in human lung cancers and enhances cell proliferation. *Cancer Research*, 65: 9628–9632.
20. Hoogsteen K (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16: 907–916. doi:10.1107/S0365110X63002437.
 21. Huang, J.C., Babak, T., Corson, T.W., Chua, G., Khan, S., Gallie, B.L., Hughes, T.R., Blencowe, B.J., Frey, B.J., Morris, Q.D. (2007). Using expression profiling data to identify human microRNA targets. *Nature Methods*, 4, 1045-1049.
 22. Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, et al. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nature Chemical Biology*, 3, 166–173. PMID:17259978.
 23. Kang, Y., Liu, Y., Wang, M.-L., Guo, M., Wang, Y., Cheng, Z.-F. (2017). Construction and analyses of the microRNA-target gene differential regulatory network in thyroid carcinoma. *PLoS ONE*, 12 (6), e0178331. <http://doi.org/10.1371/journal.pone.0178331>.
 24. Laxman, N., Rubin, C.J., Mallmin, H., Nilsson, O., Pastinen, T., Grundberg, E., Kindmark, A. (2015) Global miRNA expression and correlation with mRNA levels in primary human bone cells. *RNA*, 21, 1433–1443.
 25. Lehmann U, Hasemeier B, Christgen M, Müller M, Römermann D, Länger F et al. (2008) Epigenetic inactivation of microRNA gene hsa-miR-9-1 in human breast cancer. *Journal of Pathology*, 214: 17–24.
 26. Li, L.C., Okino, S.T., Zhao, H., Pookot, D., Place, R.F., et al. (2006) Small dsRNAs induce transcriptional activation in human cells. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 103, 17337–17342. PMID: 17085592.

27. Li, Y., Syed, J., Sugiyama, H. (2016). RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chemical Biology*, 23 (11), 1325-1333.
28. Lombe, C. (2015). Identification of microRNAs as a class of biomarkers for the early diagnosis of prostate cancer: an *in silico* and molecular approach, MSc thesis, University of the Western Cape, Cape Town.
29. Lubitz, D., Zikich, A., Kotlyar. (2010). Specific high-affinity binding of thiazole orange to triplex and G-quadruplex DNA. *Biochemistry*, 49, 3567-3574.
30. Lujambio A, Calin GA, Villanueva A, Ropero S, Sanchez-Cespedes M, Blanco D et al. (2008). A microRNA DNA methylation signature for human cancer metastasis. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 105: 13556–13561.
31. Majid, S., Dar, A. A., Saini, S., Yamamura, S., Hirata, H., Tanaka, Y., Deng, G., Dahiya, R. (2010). MicroRNA-205-directed transcriptional activation of tumour suppressor genes in prostate cancer. *Cancer*, 116 (24), 5637-49.
32. Nikolova, E. N., Zhou, H., Gottardo, F. L., Alvey, H. S., Kimsey, I. J., Al-Hashimi, H. M. (2013). A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers*, 99 (12), 955-68.
33. Nunez-Iglesias, J., Liu, C.C., Morgan, T.E., Finch, C.E, Zhou, X.J. (2010). Joint genome-wide profiling of microRNA and mRNA expression in Alzheimer's disease cortex reveals altered microRNA regulation. *PLoS One* 5, e8898.
34. Paugh, S. W., Coss, D. R., Bao, J., Lauder milk, L. T., Grace, C. R., Ferreira, A. M., ... Evans, W. E. (2016). MicroRNAs Form Triplexes with Double Stranded DNA at Sequence-Specific Binding Sites; a Eukaryotic Mechanism via which microRNAs Could Directly Alter Gene Expression. *PLoS Computational Biology*, 12 (2), e1004744. <http://doi.org/10.1371/journal.pcbi.1004744>.

35. Pfaffl, M.W, (2004). *Quantification strategies in real-time PCR*. Chapter 3, 87-112 in: A-Z of quantitative PCR (Editor: S.A. Bustin), International University Line (IUL), La Jolla, CA, USA.
36. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>
37. Tagawa, H., Seto, M. A. (2005). MicroRNA cluster as a target of genomic amplification in malignant lymphoma. *Leukemia*; **19**: 2013–2016.
38. van Der Heul-Nieuwenhuijsen, L., Dits, N.F., Jenster, G. (2009). Gene expression of forkhead transcription factors in the normal and diseased human prostate. *BJU International*, 103; 11 1574-1580, <https://doi.org/10.1111/j.1464-410X.2009.08351.x>.
39. Vasudevan S. (2012). Posttranscriptional upregulation by microRNAs. *Wiley Interdisciplinary Reviews: RNA*, 3 (3), 311–330.
40. Vasudevan, S., Tong., Y, Steitz, J.A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. *Science*, 318, 1931–1934.
41. Walz, A.L., Ooms, A., Gadd, S., Gerhard, D.S., Smith, M.A., Guidry, Auvil, J.M. (2015). Recurrent DGCR8, DROSHA, and SIX homeo-domain mutations in favourable histology Wilms tumours. *Cancer Cell*, 27, 286–297.
42. Wang, B., Hsu, S.H., Wang, X., Kutay, H., Bid, H.K., Yu, J *et al.* (2014). Reciprocal regulation of microRNA-122 and c-Myc in hepatocellular cancer: role of E2F1 and transcription factor dimerization partner 2. *Hepatology*, 59, 555–566.
43. Wang, J. (2013). Pearson Correlation Coefficient. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) *Encyclopedia of Systems Biology*. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_372.

44. Wang, L., Zhu, J., Deng, F.Y. *et al.* (2018). Correlation analyses revealed global microRNA–mRNA expression associations in human peripheral blood mononuclear cells. *Molecular Genetics Genomics* 293: 95. <https://doi.org/10.1007/s00438-017-1367-4>.
45. Wang, Y., Medvid, R., Melton, C., Jaenisch, R., & Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics*, 39 (3), 380-5.
46. Wang, Y., P, Li, K.B. (2009). Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. *BMC Genomics*, 10:218. <https://doi.org/10.1186/1471-2164-10-218>.
47. Watson, J.D, Crick, F.H. (1953). "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". *Nature*. 171 (4356): 737–738. Bibcode: 1953 doi:10.1038/171737a0. PMID: 13054692.

Chapter 3

MicroRNA and Transcription Factor gene co-regulation

3.1.1 Background

Proteins perform several specific functions within the cell. Thus, the proper functional integration and development of an organism is dependent on the manufacturing and control of these proteins within set defined concentrations, which are spatially dependent (Yilmaz and Grotewold, 2010). Protein expression can be controlled at multiple points during the central dogma of molecular biology including at the chromatin structure level, transcription, post-transcriptional processing and translation (Schwanhäusser *et al.*, 2011). Proteins can also be regulated via the action of microRNAs. Additionally, the transport and stability of a protein is also an integral part to the regulation of gene expression. Following translation and processing, proteins must be carried to their site of action in order to be biologically active. The stability of the protein varies depending on specific amino acid sequences present in the proteins as well as on its resulting folding (Hebert and Molinari, 2007; Alberts *et al.*, 2010). The primary focus of this chapter is to use computational approaches to investigate gene regulation at the transcription level of the genes identified, specifically focusing on promoters and transcription factors (TFs) and their potential role in diagnostics of prostate cancer. In particular, those TFs that regulate the genes targeted by the microRNA of interest hsa-miR-5698.

3.1.2 Regulation of Gene Expression

The regulation of gene expression is the control of the amount of a functional product of a gene as well as the timing of its appearance in a cell (Tomilin, 2008; Schwanhäusser *et al.*, 2011). This control of expression is vital because it allows a cell to produce only what it needs and where it needs it to avoid unnecessary energy expenditure and in turn allowing the cell to adapt to variable environments (Schwanhäusser *et al.*, 2011). Additionally, regulation of gene expression is the underlying basis of control of form, function and structure of an organism, particularly cellular differentiation, morphogenesis and the versatility of an organism (Yilmaz and Grotewold, 2010; Schwanhäusser *et al.*, 2011).

There are a number of regulatory mechanisms, which include those acting at the transcription level and at the translation level. At the translation level, gene regulation is carried out in various ways, including in the following mechanisms; (i) the mRNA may undergo secondary structure folding and inhibit its entrance into the ribosome, (ii) antisense RNA binding, (iii) the microRNA binding and (iv) protein binding. The transcription level involves various regulatory proteins; these include transcription factors, enhancers, activators and sigma factors. Upon binding to DNA, these regulatory proteins convey signals to the basal transcriptional machinery, containing the respective RNA polymerases, resulting in particular rates of gene expression (Boeva, 2016). Identifying and characterizing transcription factors and other DNA binding motifs that control the current genes of interest is a prerequisite to understanding the regulation and function of these genes in prostate cancer networks.

3.1.2.1 Promoters

A gene promoter is a region of DNA that initiates transcription of a particular gene (Sharan, 2007). Promoters are located near the transcription start sites of genes (TSS) or near a cluster of transcription start sites (TSSs). The role of promoters in transcriptional regulation of genes is as transcription factor binding sites (TFBSs). These are located upstream to the gene being regulated and can vary in length from 100 to 1000 base pairs (Sharan, 2007). Promoters contain specific DNA sequences that provide a secure initial binding site for RNA polymerase and for transcription factors, which are known to recruit the RNA polymerases. Since TFs have specific activator or repressor sequences that, attach to specific promoters thereby regulating gene expression (Sharan, 2007; Boeva, 2016).

3.1.2.2 Transcription factors (TFs)

The rate of transcription of genetic information from DNA to mRNA is controlled by transcription factors (TFs). Transcription factors serve as activators or repressors of transcription and a number of ways in which this is achieved have been recognized (Lambert *et al.*, 2018). As activators, transcription factors bind to promoters after which the TF interacts with components of the RNA polymerase. This interaction attracts the RNA polymerase in the vicinity of the gene promoter, thereby facilitating its binding to the core promoter. When the TF binds the DNA, the chromatin

structure in the promoter region changes and the binding area of the RNA polymerase becomes more accessible (Sharan, 2007).

As a repressor, a transcription factor will compete with an activator for a particular binding site, thereby decreasing the effect of the activator. This leads to less efficient binding of the RNA polymerase to the promoter, resulting in lower expression levels of the gene (Sharan, 2007).

3.1.2.3 MicroRNAs

MicroRNAs are naturally occurring endogenous, single stranded non-coding RNA molecules that range between 18 – 24 bases in length (Munker and Calin, 2013). MicroRNAs control gene expression by binding to target mRNAs with imperfect complementarity within the 3'-UTR, leading either to repression of translation or mRNA degradation (Cannell *et al.*, 2008). Ambros *et al.*, 1993 first discovered them in *Caenorhabditis elegans*. 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. MicroRNAs 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). Thus, dysregulation at any point in these processes, owing to microRNA mutation or abnormal expression can result in a diseased state. Therefore, this work examines the role of these non-coding RNAs and their roles in gene regulatory networks with transcription factors in prostate cancer.

3.1.3 Transcription factor and microRNA co-regulatory networks

Gene expression in cells is controlled at both the transcription and post-transcriptional levels. Transcription factors coordinate gene transcription, while microRNAs regulate gene expression by mediating post-transcriptional events, such as mRNA degradation and protein translation (Yang *et al.*, 2013). MicroRNAs and TFs co-regulate gene expression via feed-forward loops (FFLs) and feedback loops (FBLs) (Shalgi *et al.*, 2007; Yang *et al.*, 2007). In the FFL, either of two outcomes can be achieved; (i) a transcription factor will regulate a microRNA, by activating or repressing it (Figure 3.1 A). (ii) The microRNA will regulate the transcription factor by repressing it (Figure 3.1 B). Either scenario will lead to the co-regulation of their gene target (Figure 3.1 A and B) (Shalgi *et al.*, 2007; Zhang *et al.*, 2015).

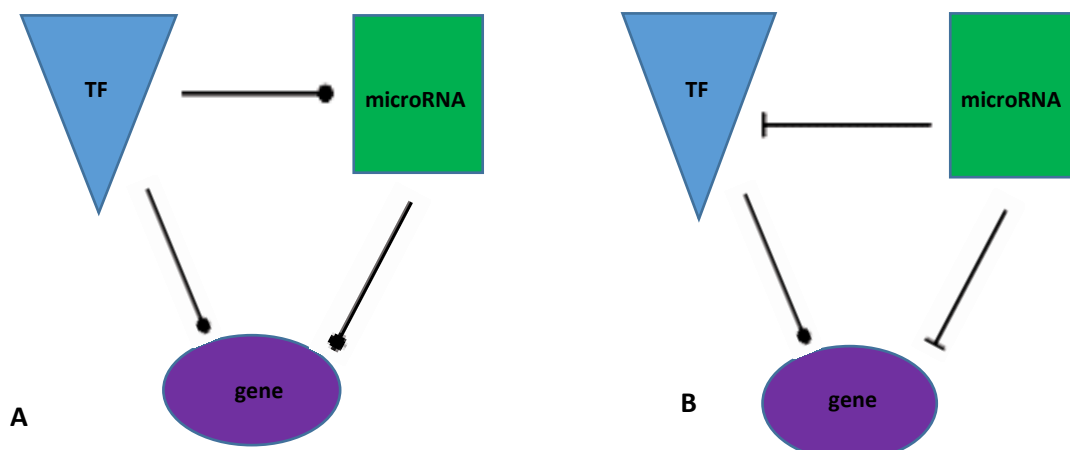


Figure 3.1: Feed Forward Loops classified by the master regulator. Blunt arrows with dot end represent transcriptional activation or repression, T-shaped arrow represents repression.

On the other hand, the feedback loop motif acts differently in that the TF and the microRNA regulate each other and each of them then regulates their targets separately. Two types of FBLs exist, the single negative and double negative loops (Figure 3.2 A and B). In the single negative feedback loop, a TF activates the transcription of a microRNA, which in turn inhibits the translation of the TF (Zhang *et al.*, 2015). An example of this is the microRNAs in the miR-17-92 cluster which participate in a single negative feedback loop with the TF E2F1 whose levels are critical for cell cycle progression (O'Donnell *et al.*, 2005; Bommer *et al.*, 2007). In the double negative loop, the TF and the microRNA repress each other (Figure 3.2 A and B). This loop exists in two steady states, where the TF is on and the microRNA is off. The double negative loop thus has key roles in cell differentiation (Zhang *et al.*, 2015).



Figure 3.2: Feedback Loops. A is the single negative loop and B is the double negative loop. The sharp arrow indicates activation.

3.1.4 DNA Sequence motifs (transcription factor binding sites)

A DNA sequence motif is a nucleotide sequence pattern that has, or is presumed to have biological significance. Sequence motifs often indicate sequence specific binding sites on DNA for proteins such as nucleases and transcription factors (TF) making them important in gene regulation (D'haeseleer, 2006). Studies have shown that the presence of certain sequences in promoter regions can determine how effective the regulation of gene transcription can be in regulatory networks and disease (Barash *et al.*, 2003; Casimiro *et al.*, 2008).

Promoter regions contain various regulatory sequences, such as the aforementioned DNA sequence motifs, which are necessary to control gene transcription (Abe and Gemmell 2014). They are part of gene regulatory networks and are known as *cis*-regulatory elements because they act near the gene of interest binding transcription factors to perform their gene regulatory functions (Wittkopp and Kalay, 2011). Promoters and their *cis*-regulatory element compositions are the initial checkpoints for transcriptional gene activities and they define the potential spatiotemporal expression of a gene (Mariño-Ramírez *et al.*, 2009). Various studies have shown that any structural changes in DNA sequence motifs can affect the *cis*-regulatory activity critically (Wray *et al.*, 2003; Vinces *et al.*, 2009; Bolton *et al.*, 2013; Abe and Gemmell 2014).

Sequence motifs can be key factors in the maintenance of promoter architecture (D'haeseleer, 2006). This is especially important when looking at regulation of genes in a particular disease such as prostate cancer networks. This chapter will investigate novel DNA sequence motifs present in the promoters of genes regulated by the microRNA hsa-miR-5698 in

prostate cancer, their role in cis-regulatory activity and consequently their biological significance and relevance in the disease.

The identification of DNA sequence motifs is a difficult problem in molecular biology. This can be alluded to the fact that a transcription factor may bind to regions that are highly variable (Hertzberg *et al.*, 2005; Hernandez-Garcia and Finer 2014). Wet bench methodologies such as the STARR-seq (self-transcribing active regulatory region sequencing) and *in vitro* translation assays have been developed to overcome this problem. However, they are very costly and time consuming. Thus, making use of the availability of complete genomic sequences including the intergenic regions, computational methods can be used to better understand these regulatory mechanisms (Hernandez-Garcia and Finer 2014).

Currently, the algorithms and tools for searching DNA sequence motifs can be divided into two major classes: Firstly, methodologies that search for well-known validated DNA sequence motifs, for example using the position specific scoring matrices (PSSMs) which is demonstrated in MatInspector, a database curating transcription factor binding sites (Quandt *et al.*, 1995; Elkon *et al.*, 2003; Aerts *et al.*, 2003). The second major class involves methodologies that try to detect novel motifs within a set of DNA sequences as demonstrated by the MEME (Multiple Em for Motif Elicitation) Suite (Bailey and Elkan 1994; Bailey *et al.*, 2006; Bailey *et al.*, 2009; Bailey *et al.*, 2009). This chapter presents a method that searches for novel transcription factor binding sites using the database MEME Suite. There are a number of bioinformatics (computational) platforms were used to perform the experiments in this chapter. An introduction of these platforms and their functions follows below.

3.1.5 Gene/protein interaction analysis using the STRING database

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. Presently, 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).

One such online database that predicts protein-protein/gene interactions is the STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins). The database is accessible at <http://string-db.org/>. This database aims to provide a comprehensive, yet quality controlled collection of protein-protein associations for a large number of organisms (von Mering *et al.*, 2003; von Mering *et al.*, 2007; Franceschini *et al.*, 2013). The protein associations are derived from high throughput experimental data, from the mining of databases and literature, and predictions based on genomic context analysis (von 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.

The STRING database currently holds 730 000 proteins in 180 fully sequenced organisms (von Mering *et al.*, 2005). The database has three unique features for protein interaction prediction. Firstly, it provides uniquely comprehensive coverage, with over 1000 organisms, 5 million proteins and more than 200 million interactions stored. Secondly, it is one of very few sites to hold experimental, predicted and inferred interactions, together with interactions obtained through text mining. Thirdly, it

includes a wealth of accessory information, such as protein domains and protein structures (Franceschini *et al.*, 2013). In this study, this database was used to find genes that may be co-expressed with the six target genes.

3.1.6 The IntAct database for co-expression analysis

The IntAct database is an open source database and software suite for modelling, storing and analysing molecular interaction data. Understanding the interactions that a protein makes with the molecules in its immediate environment, is critical for a full understanding of the processes in which that protein is involved and the mechanisms by which it is regulated (Kerrien *et al.*, 2007; Kerrien *et al.*, 2012). The IntAct database is built with data procured from published literature. The data is manually annotated to take into consideration, experimental methods used to obtain the data as well as conditions of the experiment and interacting domains of the proteins (Hermajokob *et al.*, 2004; Kerrien *et al.*, 2012). The database includes more than 126 000 binary protein interactions extracted from over 2100 scientific publications (Kerrien *et al.*, 2012). The web site provides tools allowing users to search, visualize and download data from this repository. The IntAct database is accessible at <http://www.ebi.ac.uk/intact> . It was used to find genes co-expressed with our gene target list.

3.1.7 The KEGG Pathway database

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a database that curates a wide knowledge base that can be used in the analysis of gene functions as well as the linkage of genomic information with higher order functional information in organisms (Kanehisa and Goto, 2000). The KEGG database has an array of directories embedded that store various information. The genomic information is stored in the Genes database, which is a collection of gene catalogues for all completely sequenced genomes and some partial genomes. This database also includes an up-to-date annotation of gene functions (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2016). The Pathway database stores functional information which includes the graphical representations of various cellular processes including metabolism, signal transduction, cell cycle and membrane transport. The Ligand database contains information about chemical compounds, enzymatic molecules and enzymatic reactions (Kanehisa *et al.*, 2016). The KEGG database is regularly updated and made freely available at <http://www.genome.ad.jp/kegg/>. This database was used to examine the roles of the identified target genes in prostate cancer.

3.1.8 Ensembl

The Ensembl database provides genomic information across the species chordate and some selected eukaryotes. The result of this is unique

datasets of gene annotation, multiple alignments, gene homology relationships and regulatory annotations. The database also integrates resources from other domains such as UniProt, HGNC and the Encyclopedia of DNA Elements (ENCODE) portal at UCSC. The Ensembl database is available at <http://www.ensembl.org> (Aken *et al.*, 2017). The database was used to extract promoter sequences of the target genes in this study.

3.1.9 MEME (Multiple Em for Motif Elicitation)

The MEME is a tool for discovering novel un-gapped (recurring fixed length) motifs in a group of related DNA sequences (Bailey *et al.*, 1994). The MEME tool takes a group of sequences as input and as output provides as many motifs as requested. The tool uses statistical modelling techniques to choose the best width (number of characters in the sequence pattern), number of occurrences and description for each motif pattern (Bailey *et al.*, 1994; Bailey *et al.*, 2009).

The MEME tool represents motifs as position-dependent letter-probability matrices, which describe the probability of each possible letter at each position in the pattern (Bailey *et al.*, 200). The MEME online portal is accessible at <http://meme-suite.org/tools/meme>.

3.1.9.1 TOMTOM

The TOMTOM web application found under the MEME SUITE database, compares an input DNA motif to the elements of a database of known motifs and their DNA reverse complements (Gupta *et al.*, 2007; Bailey *et al.*, 2009). For each query, TOMTOM reports a list of target motifs, ranked by a p-value. The E-value and the q-value for each match are also reported. The q-value is the minimal false discovery rate at which the observed similarity is deemed significant (Gupta *et al.*, 2007). A comparison of *de novo* motifs to databases of known motifs is important in the assignment of function to the identified motifs. For the purpose of this study, the JASPAR database was used as the search medium (Sandelin *et al.*, 2004; Mathelier *et al.*, 2016). This database curates transcription factors, and is a standalone database linked to the MEME Suite platform. The TOMTOM online portal is accessible through MEME SUITE at <http://mccb.umassmed.edu/meme/cgi-bin/tomtom.cgi>.

3.1.9.2 Gene Ontology for Motifs (GOMo)

The acronym GOMo stands for Gene Ontology for Motifs. This application is found under the MEME SUITE database, and is launched at <http://meme-suite.org/tools/gomo>.

The program performs a genome wide promoter scan using the provided nucleotide motifs as the input, to determine if any motif is significantly associated with genes linked to one or more Gene Ontology (GO) terms.

The significant GO terms can suggest the biological role/s of the motifs (Buske *et al.*, 2010). The program determines the significance of a GO term by estimating the Mann-Whitney rank-sum p-value of the term's genes. The significant GO terms associated with the target genes of the motif are sorted by a q-value. A q-value is derived from a p-value following the method of Benjamini and Hochberg, where 'q-value' is the minimal false discovery rate at which a given GO-term is deemed significant (Bailey *et al.*, 2009; Buske *et al.*, 2010).

3.1.10 Previous work

The microRNA, hsa-miR-5698 was found to be up-regulated in prostate cancer cell lines LNCaP and PC3 in a previous Master's study in the lab (Lombe, 2015). A set of bioinformatics tools was used to predict the targets of this microRNA in order to best understand its role in prostate cancer onset and progression. The microRNA was predicted to target the following genes, CDKN1A/p21, CTNND1, ELK1 BIRC2, FOXC1 and LRP8. A literature review on these genes indicated that p21 and BIRC2, FOXC1 and CTNND1 are up-regulated in prostate cancer with p21 being a major inhibitor of tumour cell proliferation. The genes ELK1 and LRP8 are recorded to be down-regulated in PCA.

Aim

As mentioned earlier, because gene expression relies on both transcriptional and translational control, this chapter aims to, Identify novel transcription factor binding sites in the promoters of hsa-miR-5698 target genes and investigate their biological significance in the progression of PCa.

Specific objectives:

1. To identify genes that are co-expressed with the hsa-miR-5698 target genes as co-expressed genes are potentially co-regulated using STRING.
2. Understand role of co-expressed genes in PCa by pathway analysis in KEGG.

3. Identify novel putative transcription factor binding sites (motifs) in the promoters of the target genes of hsa-miR-5698 and their co-expressed genes using MEME Suite.
4. Identify the putative transcription factors that bind to the motifs using JASPAR database.
5. Use databases and literature mining using tools such as Gene Expression Atlas, Protein Expression Atlas, The Cancer Genome Atlas and Human Proteinpedia Database (HPRD) to determine the expression profiles of the identified transcription factors in prostate cancer cell lines, and tissue.
6. Use obtained information and correlation analysis to build putative hsa-miR-5698 and TF gene regulatory networks involved in PCa progression.

3.2 Methodology

3.2.1 (a) Co-expression analysis via STRING

The official gene IDs for the six genes (**ELK1**, **LRP8**, **BIRC2**, **CTNND1**, **CDKN1A**, **FOXC1**), targeted by hsa-miR-5698 were used as input for the generation of a gene network using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database Version 10.5 (Franceschini *et al.*, 2013; Szklarczy *et al.*, 2017). The STRING database is launched at <https://string-db.org/>. The combined set of six genes was used as a driver to produce a co-expression network. Co-expressed genes are known to be co-regulated (Meier 2011) and thus their upstream sequences may share common DNA motifs.

To generate a co-expression network in STRING, the official gene identifiers of the six targets were pasted into the search box and *Homo sapiens* was selected as the organism. Additionally, the following parameters in the STRING database were chosen; (i) a confidence level of 0.7, (ii) a network depth of four and (iii) restriction to show only the top 50 interactions between the six genes targeted by miR-5 and their co-expressed genes.

3.2.1 (b) Co-expression analysis via IntAct

The official gene IDs for the six genes, targeted by hsa-miR-5698 were used as input for the generation of a gene network in the IntAct Molecular Interaction Database curated by the European Bioinformatics Institute

available at <https://www.ebi.ac.uk/intact/>. The number of interactions were downloaded in table form and Cytoscape 3.5.1 (Shannon *et al.*, 2003) was used to visualize the interactions. Cytoscape is an open source platform for complex network data Integration, analysis, and visualization. It is available for download at <https://cytoscape.org/>. A final working gene list of the six genes targeted by hsa-miR-5698 and genes common to both the STRING and IntAct co-expression networks were selected for further analysis. The total number of genes in the generated co-expressed network was 11.

3.2.2 Extraction of gene promoters from the Ensembl database

Promoters sequences were extracted for the set of 11 genes using the Ensembl Genome Browser 91 and saved in FASTA format. The promoter sequences 1000 bp upstream and 200 bp downstream from the transcription start site were extracted in Ensembl by pasting the official gene ID into the search box provided on the website and pressing the return button. The configure settings in Ensembl were used to adjust the flanking positions around the transcription start site. Each of the retrieved promoters was 1000 bp long.

To make certain that the promoter sequences obtained from Ensembl were bona fide, the BLAT tool of the UCSC Genome Browser (accessed at <http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) was used to align them to the sequences annotated in the UCSC database and checking for

matches. The BLAT tool also has a CpG islands feature that was utilized to further verify that the promoter sequences were true.

3.2.3 *De novo* motif discovery using MEME (Multiple Em for Motif Elicitation) SUITE

The MEME Suite online portal accessed from <http://meme-suite.org/tools/meme> was used to submit the unaligned promoter sequences of the 11 genes. The following parameters were used in the purpose of this analysis; DNA was selected for the sequence alphabet, motif discovery mode was set to normal, site distribution was selected to zero or one occurrence per sequence and number of motifs was set to 10. The rest of the parameters of the database were left at default. The output from MEME was appropriately saved for further analysis.

3.2.4 Comparison of discovered motifs to elements of databases of known motifs using TOMTOM

The identified set of *de novo* motif sequences from MEME were used as the input into TOMTOM, which is accessed at <http://mccb.umassmed.edu/meme/cgi-bin/tomtom.cgi>. The parameters were set as follows; search database, JASPAR CORE and the motif column comparison function was set to the Pearson correlation coefficient. The rest of the parameters were left at the default setting.

3.2.5 Gene Ontology (GO) analysis for the discovered motifs using Gene Ontology for Motifs (GOMo)

The identified set of *de novo* motif sequences from MEME were used as the input into GOMo, which is accessible at <http://meme-suite.org/tools/gomo>. The application has one parameter, which was set to search promoters of *Homo sapiens*.

3.2.6 Building the regulatory relationships among hsa-miR-5698, its target genes and putative transcription factors in PCa

The computational targets of the microRNA hsa-miR-5698 were predicted in the TargetScanHuman, MAGIA and miRBase databases in a Master's study (Lombe, 2015). The TargetScan database and the MATCH tool of the Transfac database were used to identify hsa-miR-5698 and transcription factor regulatory pairs. The Transfac database is launched from <http://gene-regulation.com/>.

The expression data generated via qPCR in Figure 5.1 as well as the Pearson Correlation coefficient calculated in Chapter 2 and shown in Table 2.1 were used to select significant relationships to build networks. A correlation coefficient ≤ -0.20 with a p-value < 0.05 for all pairs. The relationship pairs investigated were, hsa-miR-5698-mRNA, hsa-miR-5698-TF, TF-gene and TF-hsa-miR-5698. Literature mining was undertaken on the putative transcription factors linked to motif 1 and motif 2 and their expression profiles in prostate cancer. The data obtained was used in the construction of the hsa-miR-5698-TF, TF-gene and TF-hsa-miR-5698 regulatory relationships.

3.3 Results and discussion

3.3.1 Co-expression analysis via STRING and IntAct

When genes share a similar expression profile under specific conditions such as a particular disease like prostate cancer, they may under-go co-expression. Co-expression may imply the presence of a functional biological linkage between the genes such as carcinogenesis (Tavazoie *et al.*, 1999). Thus, a set of prostate cancer biomarkers can be used in the generation of co-expression networks to identify other markers that can predict disease stage as well as the clinical outcome for the disease. This could help understand the disease pathogenesis and provide personalized treatment (Yuan *et al.*, 2017).

One important goal of analysing gene expression data is to discover co-regulated genes. Similar patterns in gene expression profiles usually suggest relationships between the genes. According to (Lee *et al.*, 2002), genes targeted by the same transcription factors tend to show similar expression patterns spatially.

Co-expression analysis was performed by the STRING and IntAct databases using the six targets of hsa-miR-5698 as input. The result from both databases was an intersection set of genes co-expressed with the six input genes. The intersection set was **PCNA, CDK4, CDK2, TP53** and **CCND1**. A co-expression network consisting of the intersection set and the hsa-miR-5698 target genes was generated using the Cytoscape platform (Figure 3.3). There were a total of 11 genes in the co-expression network; namely, **ELK1, LRP8, BIRC2, CTNND1, CDKN1A, FOXC1, PCNA, CDK4, CDK2, TP53** and **CCND1**.

The hsa-miR-5698 targets and the co-expressed genes all showed direct interaction with each other (Figure 3.3). The gene CDKN1A (p21) is a central member in the co-expression network. It is a member of the cyclin-dependent kinase inhibitor family and is a major transcriptional target of the p53 protein. In prostate cancer, it inhibits cellular proliferation by inhibiting the proliferating cell nuclear antigen (PCNA); also shown in Figure 3.3. The protein product of PCNA is required for S phase progression during the cell cycle (Abbas and Dutta, 2009; Jain *et al.*, 2013). Pathway analysis via KEGG was performed on the genes in the co-expression network to better understand their roles in the processes of apoptosis, cell cycle and autophagy (three hallmarks of cancer), the results are shown in Figure 3.3.

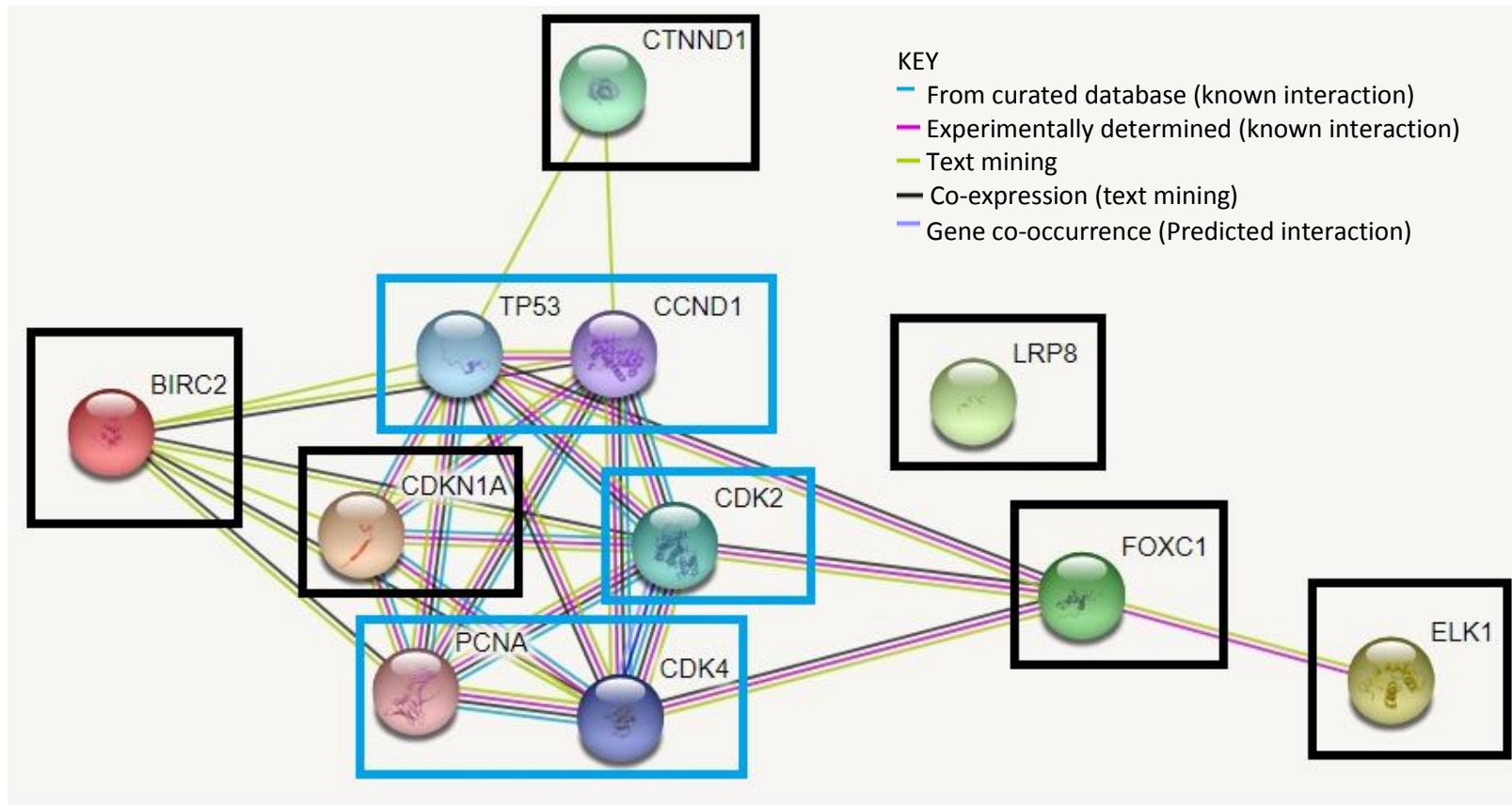


Figure 3.3: Protein co-regulation network from STRING and IntAct, visualized in Cytoscape. The microRNA gene targets are outlined in the black boxes and the co-regulated genes are outlined in blue.

3.3.2 KEGG pathway analysis

The analysis of the set of 11 genes in the KEGG pathway showed that three of the genes namely, CDKN1A/p21, CDK2 and TP53/p53 annotated by the red boxes are involved in the prostate cancer pathway (Figure 3.4). The gene CDKN1A (p21) was found to be involved the Phosphatidylinositol-3 kinase/Protein kinase B (PI3K/Akt) pathway in the KEGG database. The PI3-kinases have been linked to various cellular functions including cell growth, proliferation, apoptosis, differentiation, survival and cellular trafficking (Osaki and Oshimura, 2004).

The role of Akt in cell proliferation is as a modulator of various substrates such as the cyclin-dependent kinase inhibitors p21/Waf1/Cip1 and p27/Kip2 (Liang *et al.*, 2002; Viglietto *et al.*, 2002). The CDKN1A gene is a cyclin-dependent kinase inhibitor and encodes the p21 protein (Eastham *et al.*, 1995; Warfel and El-Deiry 2013). This protein is an inhibitor of cell cycle progression in the G1/S and G2/M transitions. It does this by inhibiting CDK4 and CDK2 respectively. The protein disrupts the phosphorylation of Rb by CDK-Cyclin, which is a regulator of cell cycle progression (Wang *et al.*, 2011). Upregulated expressions of CDKN1A by adenoviral vectors has been shown to inhibit prostate cancer growth both *in vivo* and *in vitro*. Thus, this gene is an ideal target for gain-of-function manipulation to suppress the growth of prostate cancer cells (Warfel and El-Deiry 2013). According to a previous study conducted in our laboratory, this gene is targeted by hsa-miR-5698, which the study showed to be overexpressed in prostate cancer cell lines LNCaP and PC3 using qPCR (Lombe, 2015).

Several studies have been conducted on the effect of various microRNAs and their effect on p21 in several cancers. One such study by Chuanchang

and co-workers (2017) found that p21 is a target of the microRNAs hsa-miR-1236-3p and hsa-miR-370-5p. Their study showed that these microRNAs activate p21 expression in the bladder cancer cell line T24 and the non-small-cell lung carcinoma cell line A549. However, both microRNAs failed to induce expression of p21 in the hepatocellular cell line HepG2. Similarly, a study by Li *et al* in 2017, showed transfection of hsa-miR-3617-5p into prostate cancer cell lines DU145 and PC3 induces expression of p21 by targeting its promoter region and inhibiting prostate cancer cell proliferation (Li *et al.*, 2017). From Chapter 2, the Pearson Correlation analysis revealed that hsa-miR-5698 may work to down-regulate p21 in LNCaP, but up-regulate it in PC3 as evidenced by the correlation coefficients (Figure 2.4).

Another cyclin-dependent kinase that plays an essential role in the regulation of multiple events of the cell cycle is CDK2 (cyclin-dependent kinase 2/Cyclin E). Various studies have shown that CDK2 is functionally associated with hyper-proliferation in multiple cancers and tumour cells. Thus, it could also be regarded as a potential prognostic and therapeutic target (Flores *et al.*, 2010; Chohan *et al.*, 2015). The KEGG Pathway shows that this gene is inhibited by p27 in the prostate cancer pathway (Figure 3.4). It can also be seen that when expressed, CDK2 phosphorylates the retinoblastoma protein (pRb) thereby inhibiting it from binding the transcription factor E2F, thus allowing transition of the cell from G1 to the S phase and facilitating cell growth. Additionally, CDK2 may also regulate G2 transition into M phase independently of p53, another gene that was part of our generated co-expression network using Cytoscape (Figure 3.3). This is particularly important, because there is a G2/M checkpoint that remains intact in p53- deficient cancer cells (Chung and Bunz 2010).

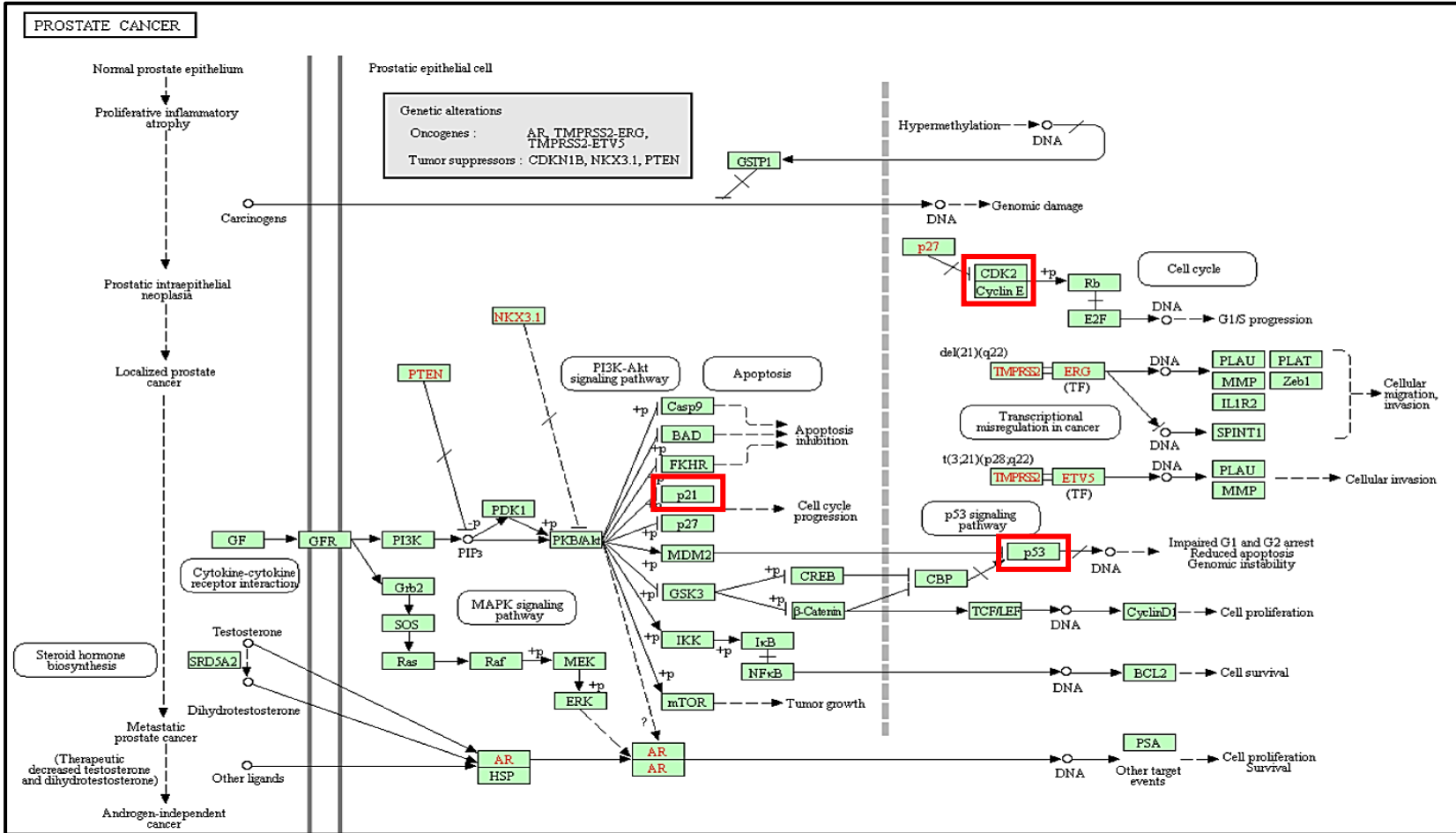


Figure 3.4: The prostate cancer pathway from KEGG database. The pathway shows the position of three of the genes of interest in the pathway (red boxes).

3.3.3. *De novo* motif discovery

Sequence motifs are short, recurring patterns in DNA that are presumed to have a biological function. Often they indicate sequence-specific binding sites for proteins such as nucleases and transcription factors (D'haeseleer, 2006). The TATA promoter sequence is an example of a highly conserved DNA sequence motif found in eukaryotes. Previously, binding sites were determined through DNase foot-printing, and gel-shift or reporter construct assays (D'haeseleer, 2006; Maclsaac, 2006). However, as stated earlier, these methods have been known to be time consuming and expensive. Currently, computational methods are generating a flood of putative regulatory sequence motifs. The common methodology for this is using algorithms to search for overrepresented (and/or conserved) DNA patterns in promoter regions of functionally related genes such as those with similar expression patterns or similar functional annotation (D'haeseleer, 2006).

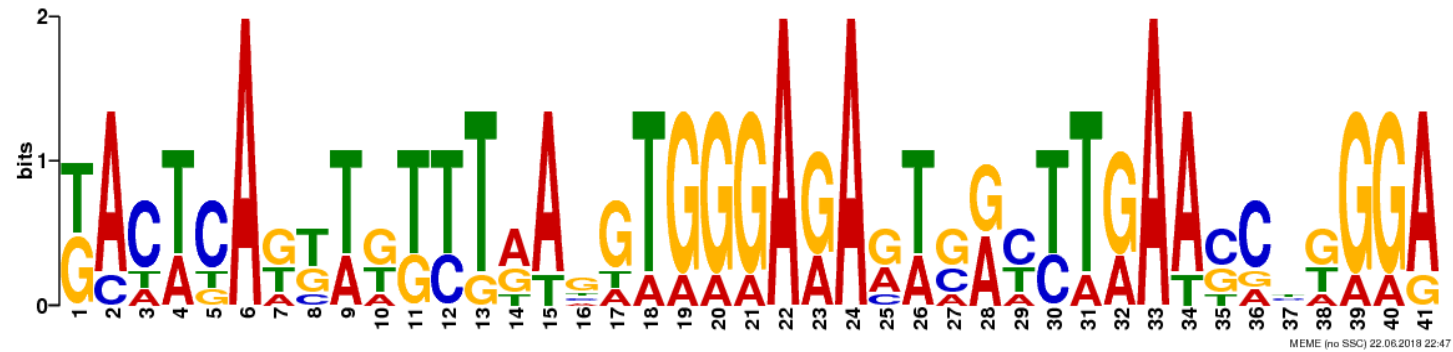
In this study, three *de novo* motifs were identified among the proximal promoter regions of the microRNA target genes and their co-regulated genes via MEME. Figure 3.5 represents a sequence logo of the discovered motifs, whilst Figure 3.6 represents their reverse complements (antisense strand). Each letter shows the conservation of that base and its frequency of occurrence at that position. In motif one, at position one, adenine has a score of 2 bits, which shows that it is conserved at that position and occurs there 100 % of the time. A number of sites contribute to the construction of each motif. For motif one, there are eight contributing sites, motif two has five and motif three has seven.

MEME reports an E-value for each motif it finds. The E-value of a motif is an estimate of the number of motifs expected to be found by chance if the letters in the input sequences were shuffled. It is based on the size of the motif, the number of sites used to build the motif and the frequency of returning the same motif over background noise upon a repeat search (Bailey *et al.*, 2006; Bailey *et al.*, 2009). The associated E-value for each motif was less than 0.05. This indicates that each motif is statically significant and is unlikely to be a random sequence artefact.

Motif 1 4.3e-004
8 sites



Motif 2 9.5e-003
6 sites

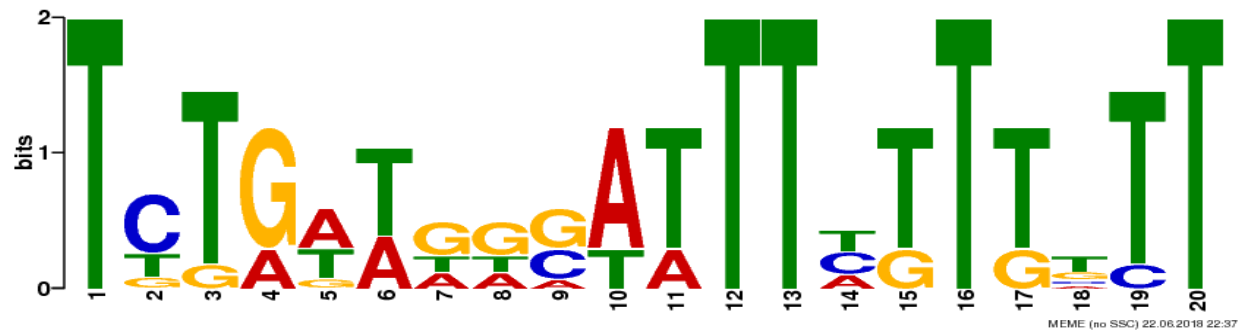


Motif 3 2.5e-001
11 sites

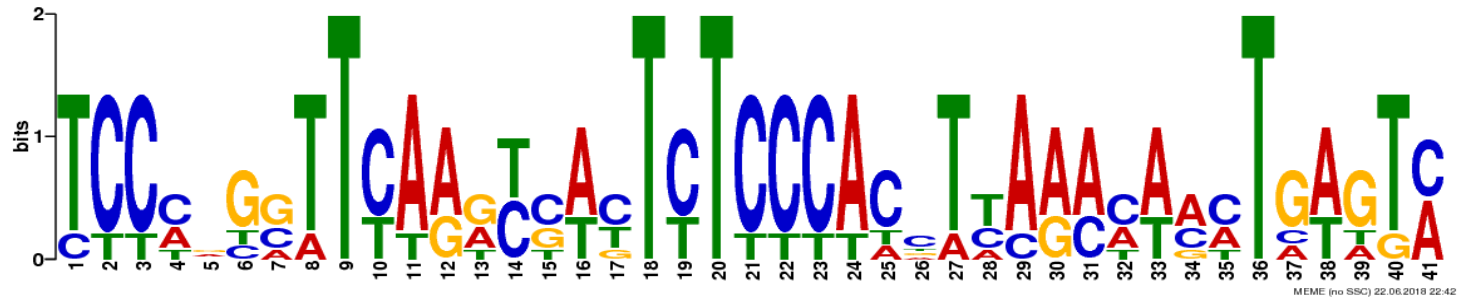


Figure 3.5: Sequence logos of three identified motifs common among the microRNA target genes and their co-regulated genes. Column one represents the motif number as designated by MEME. Column two represents the E-value associated with the motif prediction and the number of sites contributing to the construction of the motif. Column three represents the sequence logo of the motif.

Motif 1 4.3e-004
Reverse complement 8 sites



Motif 2 9.5 e-003
Reverse complement 6 sites



Motif 3 2.5.0e-001
Reverse complement 11 sites

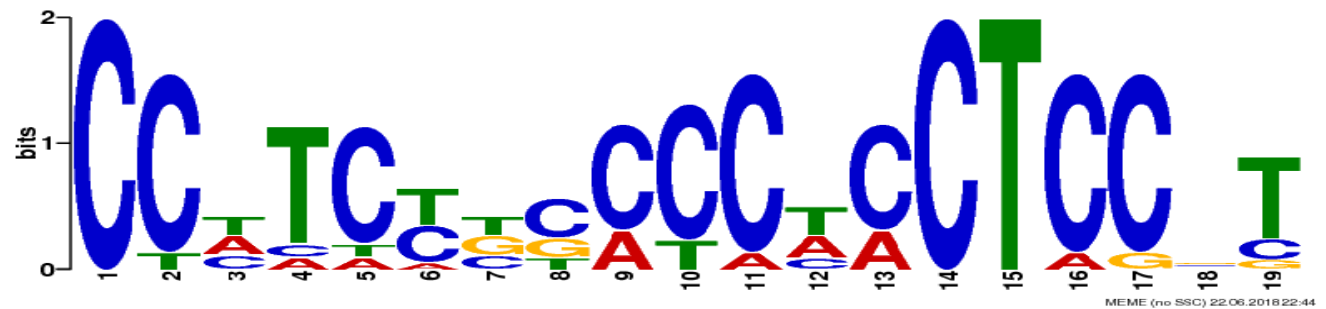


Figure 3.6: Sequence logos of the reverse complements (antisense strand) of the three identified motifs common among the microRNA target genes and their co-regulated genes. Column one represents the motif number as designated by MEME. Column two represents the E-value associated with the motif prediction and the number of sites contributing to the construction of the motif. Column three represents the sequence logo of the motif.

Figure 3.7 shows the location of each of the three motifs on the promoter sequence of each gene provided as input. Each block shows the position and strength of the motif site. The height of a block gives an indication of the significance of the site, as taller blocks are more significant. The height is calculated to be proportional to the negative logarithm of the p-value of the site. Because DNA is double stranded, the sites on the coding strand are shown above the line and those shown below the line are on the reverse complement or antisense strand (Bailey and Elkan, 1994).

From Figure 3.7, it can be seen that motif two has a greater height than motifs one and three. The figure also shows that the motif is located in various positions along the promoters of the following genes, CDKN1A/p21, CTNND1, CCND1, TP53, CDK2 and CDK4. The motif is located on the antisense strand of the promoter of genes CDKN1A, TP53, CCND1 and CDK4 and the sense strand of promoters of genes CTNND1 and CDK2. However, this does not affect the function of the motif. One study found that for many motifs, function does not depend on the orientation of binding, but rather the event of binding itself, gene-locus and specific additional factors such as enhancers or repressors (Lis and Walther, 2016).

The size of the motif is 40 bases long (Figure 3.5, 3.6 and Table 3.1). For the CDKN1A/p21 promoter, the motif begins at position 9 and ends at position 49. This is upstream of the transcription start site. It can be seen from Figure 3.7 and Table 3.1 that motif one is located up stream of the transcription start site in all the promoters it has been discovered in. Thus, this motif perhaps follows canonical transcription factor binding on the promoter sequences. Motif 3 was found in the promoters of all the genes. This coupled with its shortness in height (shortest of the three) suggests it is not a strong motif and perhaps this accounts for its commonality.

Table 3.1: Position and size of motif 2 on the promoters of the corresponding genes

Gene	Location		Width
	Starting base	Ending base	
CDKN1A/p21	9	49	40
CTNND1	120	160	40
TP53	348	388	40
CCND1	23	63	40
CDK2	467	502	40
CDK4	103	143	40

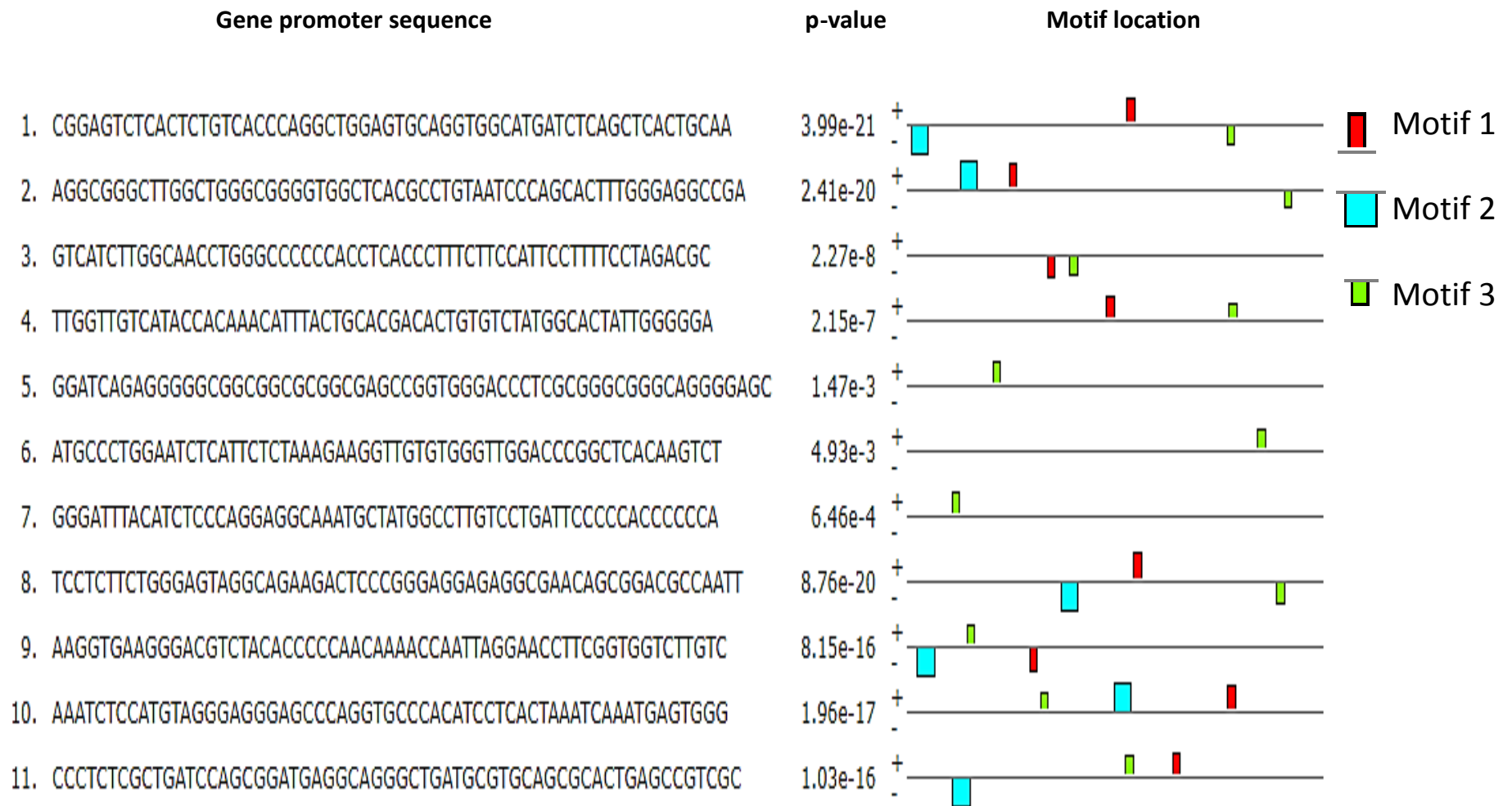


Figure 3.7: Location and p-value of the three identified *de novo* motifs on the promoters of the submitted genes. From one to 11, the gene promoters' sequences are as follows, CDKN1A/p21, CTNND1, ELK1, BIRC2, FOXC1, LRP8, PCNA, TP53, CCND1, CDK2 and CDK4. The positive sign denotes the sense strand and the negative sign denotes the antisense strand.

All three motifs were identified in the promoter of the cyclin-dependent kinase inhibitor, CDKN1A (p21) Figure 3.8. Motif 1 was identified in the promoters of two targets (ELK1 and BIRC2) of the microRNA, hsa-miR-5698 whilst motif 2 was identified on CTNND1 and motif 3 on LRP8 and FOXC1, which are also targets of the same microRNA. Among the co-expressed genes, PCNA was the only gene in which only one motif was identified (motif 3).

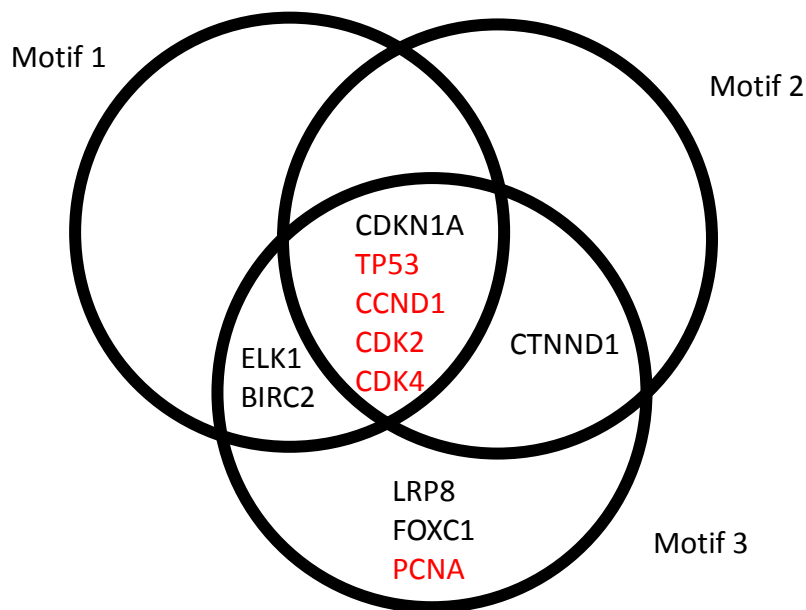


Figure 3.8: Venn diagram of motif distribution on the gene promoters. The genes in black indicate the starting gene list. These are the targets of the microRNA hsa-miR-5698. Those in red indicate the genes co-expressed (STRING/IntAct) with the hsa-miR-5698 targets.

3.3.4 Comparing the discovered motifs to known transcription factor binding profiles in the JASPAR database using TOMTOM

The use of computational methods to discover novel transcription factor binding sequence motifs (DNA sequence motifs) in a set of promoter sequences has some limitations and thus can be a difficult task. This is owing to the fact that these binding sites tend to be short and degenerate especially in eukaryotes. Additionally, searching for sequences without prior knowledge of binding sites or nucleotide patterns can be biased and validation should be performed. Thus, the MEME Suite application TOMTOM was used to compare the *de novo* motifs to the JASPAR database, which is a database of transcription factor binding profiles (Table 3.2). The TOMTOM result showed that motif 1 matched 28 transcription factor binding profiles in the JASPAR database, motif 2 matched 15 and motif 3 matched 43 profiles. Only the top five matched profiles based on statistical significance are tabulated in Table 3.2 and were examined further.

Table 3.2: Top matched transcription profiles for each *de novo* motif.

Motif	Matched profiles	Top five matched profiles	p-value	q-value	E-value
1	28	MA0041.1 (FOXD3)	5.31e-06	1.66e-02	8.30e-03
		MA1184.1 (RVE1)	6.08e-04	4.74e-01	9.50e-01
		MA0940.1 (AP1)	8.80e-04	4.78e-01	1.38e+00
		MA1125.1 (ZNF384)	9.79e-04	4.78e-01	1.53e+00
		MA0277.1 (AZF1)	1.15e-03	4.78e-01	1.79e+00
2	15	MA0688.1 (TBX2)	1.34e-03	9.39e-01	1.88e+00
		MA0802.1 (TBR1)	1.67e-03	9.39e+00	2.35e-01
		MA0800.1 (EOMES)	1.77e-03	9.39e-01	2.49e+00
		MA0803.1 (TBX15)	1.92e-03	9.39e-01	2.70e+00
		MA0685.1 (SP4)	2.06e-03	9.39e-01	2.89e+00
3	43	MA0528.1 (ZNF263)	3.98e-06	1.01e-03	6.67e-04
		MA1274.1 (OBP3)	7.26e-07	1.01e-03	1.02e-03
		MA1268.1 (AT1G69570)	1.18e-06	1.09e-03	1.66e-03
		MA1277.1 (Adof1)	1.79e-06	1.24e-03	2.51e-03
		MA1267.1 (AT5G66940)	3.02e-06	1.67e-03	4.24e-03

In the JASPAR database, one of the transcription binding profiles that was found to be structurally similar to motif 1 was MA0041.1, which is the binding profile of the TF FOXD3. This match was over the first 12 base pairs of motif 1 (Figure 3.9.) with an offset of two bases at position 5 and 8. The match was found in the *Rattus norvegicus* species database. This transcription factor is from the class of transcription factors called the Fork head or the winged helix factors. FOXD3 is a transcriptional regulator in embryogenesis and is understood to be critical for maintenance of self-renewal, survival, and pluripotency in murine embryonic stem cells (Wang *et al.*, 2016). In some cancers (colon, lung and colorectal), it acts as a

tumour suppressor and inhibits cell growth, invasion, angiogenesis and proliferation (Yan *et al.*, 2015; He *et al.*, 2016; Li *et al.*, 2017). This finding could suggest that motif 1 is a novel motif in humans and may perform similar functions.

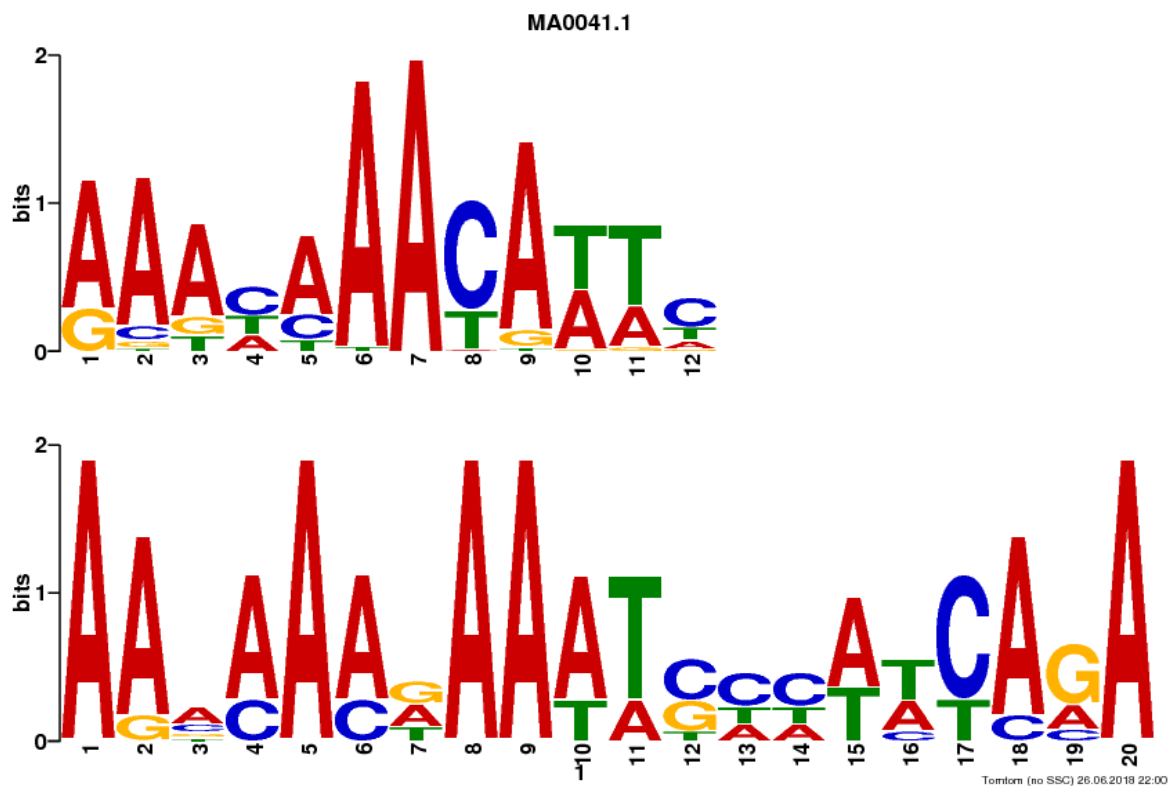


Figure 3.9: Optimal alignment of the two motifs. The sequence logo of the target motif is shown aligned above the logo for the query, which was motif 1.

Motif 2 was found to be structurally similar to the transcription binding profile, MA0688.1, which is the binding motif for the transcription factor (TBX2) a T-box transcription factor. The match was over bases 14 to 24 of motif 2, with no offsets. Thus, this may not be a novel motif. The matched transcription factor TBX2 was from the *Homo sapiens* species database. This transcription factor has been reported to be implicated in tissue development at different sites in the TiGER (Tissue-specific Gene Expression and Regulation) Database (Liu *et al.*, 2008). Aberrant expression

of the factor in many malignant tumours has been demonstrated to be conducive to tumour progression. In breast cancer, overexpression of TBX2 contributes to carcinogenesis by accelerating cell proliferation and making cells resistant to chemotherapy (Sunwoo and Suresh, 2013, p. 989). In prostate cancer, the knockdown of TBX2 in PC3 and LNCaP (prostate carcinoma and metastatic cell lines respectively) inhibits cell metastatic abilities by down regulating fibronectin, which protects prostate cancer tumours from necrosis. Subsequently, immunohistochemistry results indicate that expression levels of this factor in prostate tumour tissue are elevated when compared to adjacent normal tissues (Du *et al.*, 2017).

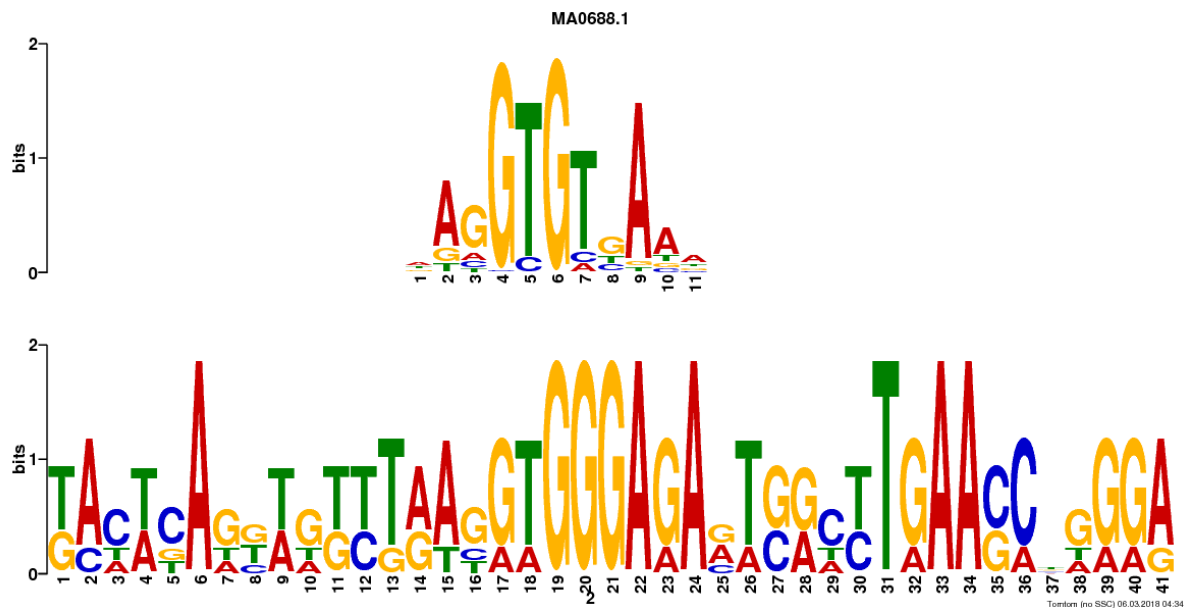


Figure 3.10: Optimal alignment of the two motifs. The sequence logo of the target motif is shown aligned above the logo for the query, which was motif 2.

The JASPAR database matched motif 3 to the transcription profile MA0528.1 which is the binding motif for the transcription factor ZNF263, Figure 3.11. It belongs to the zinc finger factors. The match was over the first 12 bases of the motif, with an offset of two bases at position 2 and 6. The matched transcription factor, ZNF263 was from the *Homo sapiens* species database.

The expression profiles of the three putative transcription factors FOXD3, TBX2 and ZNF263 were investigated in the TiGER database (Liu *et al.*, 2008). It was found that they are all expressed in prostate cancer and motif 1 was detected as a cis regulatory module in prostate cancer Figure 3.11.

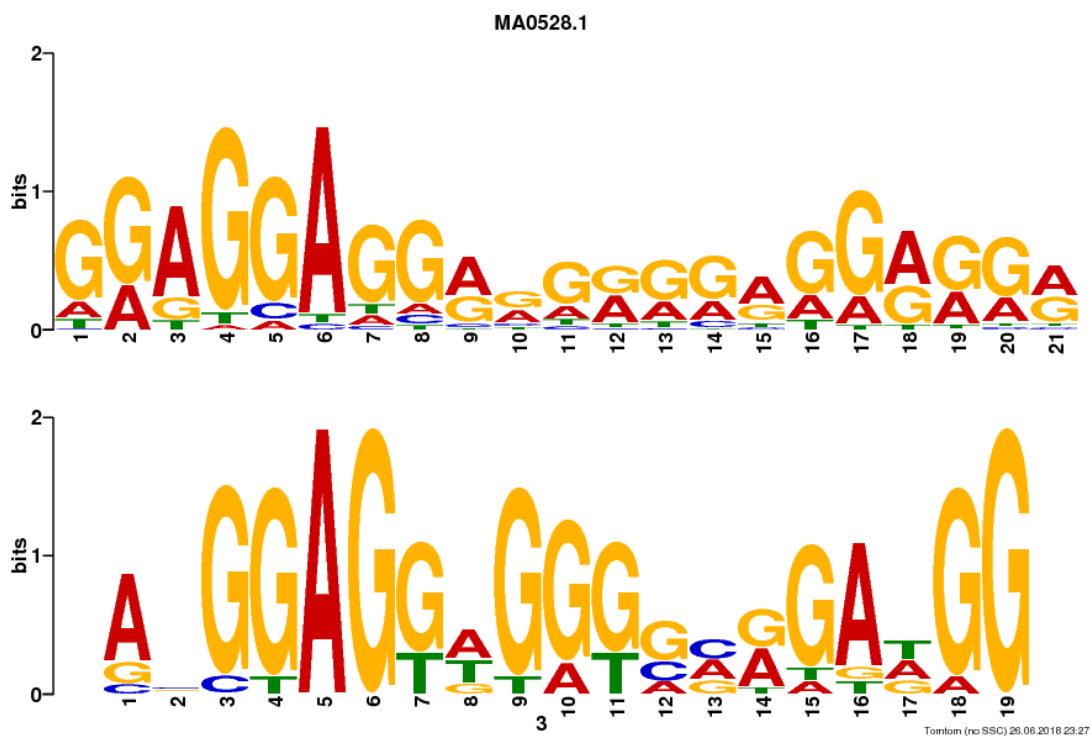


Figure 3.11: Optimal alignment of the two motifs. The sequence logo of the target motif is shown aligned above the logo for the query, which was motif 3.

3.3.5 Gene Ontology (GO) analysis via GOMo

To determine if our predicted motifs are valid and to determine their possible functions, the motifs were scanned against Gene Ontology terms using the GOMo tool in the MEME database. The gene ontology is a description of vocabularies to describe key domains of molecular biology including gene product attributes and biological sequences (Bodén and Bailey 2008; Buske *et al.*, 2010). In this case, we applied a gene ontology search on the *de novo* motifs to determine if there was any match to one or more gene ontology terms, the results of this search are tabulated in Table 3.3. All the motifs were found to be involved in the biological process of DNA damage checkpoint and repair. Additionally, motif 1 and 2 were also involved in G-protein coupled receptor activity.

Table 3.3: Representation of the total number of GO terms predicted per motif as well as the top predicted GO terms of the three motifs with BP being Biological Process, MF Molecular Function and CC cellular Component.

Motif	Total Predictions	Identified Gene ontology term (Top predictions)
1	20	BP - G-protein coupled receptor protein signalling pathway - double-strand break repair - DNA damage checkpoint - DNA integrity checkpoint MF - G-protein coupled receptor activity
2	88	CC - external side of plasma membrane MF - transmembrane receptor activity - G-protein coupled receptor activity BP - immune response, DNA damage checkpoint
3	30	BP - DNA damage checkpoint - regulation of organ growth - G-protein coupled receptor protein signalling pathway - gene expression MF - double stranded RNA binding

3.3.6 MicroRNA and DNA sequence motifs co-regulation of gene expression

As discussed in section 3.1.4, DNA sequence motifs have biological significance and often indicate sequence specific binding sites for proteins such as nucleases and transcription factors (TF). Three putative transcription factors identified as possible targets of the discovered motifs

were FOXD3, TBX2 and ZNF263, corresponding to motif 1,2 and 3 respectively. These transcription factors were used to analyse the microRNA-TF gene expression regulation networks and thus are referred to as motif-TF as described in Section 3.2.6. The Cytoscape program version 3.5.1 was used to build the interaction network between the microRNA, target genes and the desired *de novo* motif Transcription factor. Four types of regulatory relationships among genes, miRNAs and TFs were predicted. The prediction results of the regulatory relationships are summarized in Table 3.4 below. By merging the regulatory relationships predicted in table 3.4, 3-node feed forward loops (FFL) were formed, Figures 3.12 to 3.15.

Table 3.4: Summary of relationships among PCa-related genes, hsa-miR-5698, and putative TFs.

Relationship	Gene or TF match	Prediction method
hsa-miR-5698-gene	CDKN1A, CTNND1, BIRC2 FOXC1, LRP8, ELK1	TargeScanHuman, MAGIA and miRBase
hsa-miR-5698-TF	TBX2	TargeScanHuman and Transfac
TF-gene	CDKN1A, CTNND1, BIRC2 FOXC1, LRP8, ELK1, CCND1, TP53, PCNA, CDK2, CDK4	MEME Suite, JASPAR
TF- hsa-miR-5698	TBX2, FOXD3	TargetScanHuman, Transfac (MATCH™)

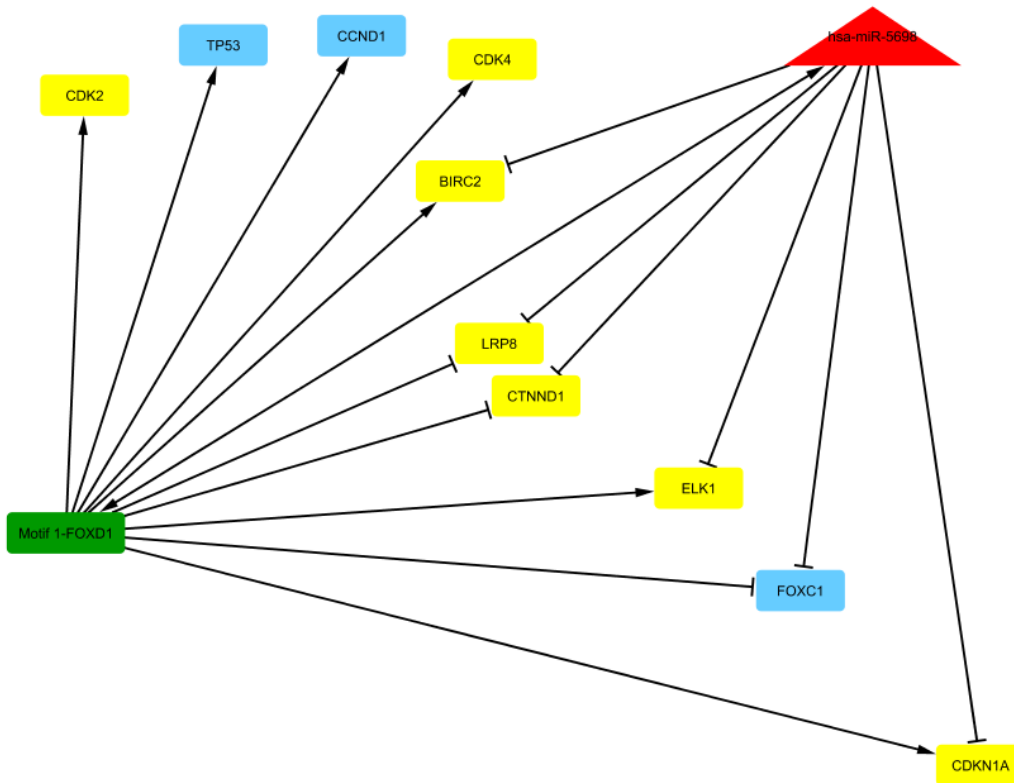


Figure 3.12: Integrative microRNA target genes and the DNA sequence motif 1 transcription factor FOXD3 regulatory networks in the prostate adenocarcinoma cell line LNCaP. The Arrows indicate activation of the target by the transcription factor and/or the microRNA. The blunt lines indicate repression of the target by the transcription factor and/or the microRNA. Both the transcription factor and the microRNA activate each other. The genes with the blue coloured nodes are down-regulated in the LNCaP cell line, whilst the ones in yellow are up-regulated.

Examining separate nodes in Figures 3.12-3.15 identifies two FFLs. The first is the microRNA-FFL, in which hsa-miR-5698 acting repressing or activating both the gene and the TF. The second is the TF-FFL, in which the TF regulates the genes as well as hsa-miR-5698. When the regulatory network

of the gene CDKN1A is taken into consideration (Figure 3.12), it can be seen that in one FFL (microRNA-FFL), the gene is being repressed in the LNCaP cell line. In the other FFL (TF-FFL), the gene is activated by FOXD1. This transcription factor is a tumour suppressor in hepatocellular carcinoma (He *et al.*, 2015), colon cancer (Li *et al.*, 2017) and thyroid cancer progression (Yin *et al.*, 2017). This may explain why the transcription factor acts against the microRNA and promotes CDKN1A whose protein p21 is an inhibitor of cell cycle progression in the G1/S and G2/M transitions (Wang *et al.*, 2011). Activation of this gene has been linked to inhibition of prostate cancer cells (Jain *et al.*, 2013; Li *et al.*, 2017).

In contrast, both the microRNA and the TF inhibit the FOXC1 gene in putative regulatory network (Figure 3.12). This gene (FOXC1) has been found to have dual effects as both a tumour suppressor and pro-metastatic mediator. Aberrant expressions of this gene have been linked to malignancy, proliferation, differentiation, survival and metastasis (Yang *et al.*, 2017; Elian *et al.*, 2018). However, overexpression of FOXC1 has also been linked to inhibition of invasive progression in prostate (van Der Heul-Nieuwenhuijsen *et al.*, 2009; Yang *et al.*, 2017). Expression profiling via qPCR conducted in LNCaP on this gene showed that it is down-regulated in the cell line (Figure 5.1). This could explain inhibition of this gene by the TF, which is a tumour suppressor.

Figure 3.1.3 represents the regulatory networks of the TF FOXD3, genes and microRNA in PC3 cell lines. A comprehensive literature mining did not yield any expression data for this transcription factor in PC3 or metastasis prostate tumour samples. Thus, this is a very weak representation of a putative regulatory network in PCa.

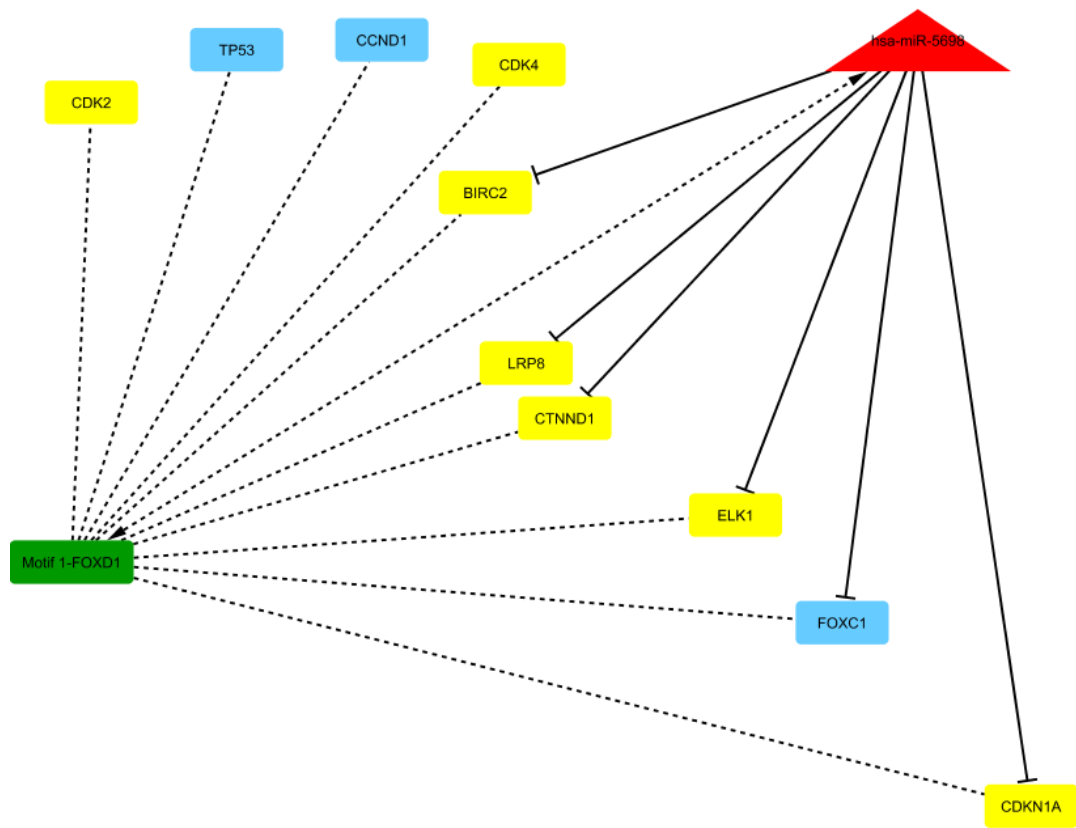


Figure 3.13: Integrative microRNA target genes and the DNA sequence motif transcription factor FOXD3 regulatory networks in the prostate metastatic cell line PC3. The blunt lines indicate repression of the target by the microRNA. The dashed lines indicate that there no data available currently to make significant predictions about the relationship. The genes with the blue coloured nodes are up-regulated in the PC3 cell line, whilst the ones in yellow are down-regulated.

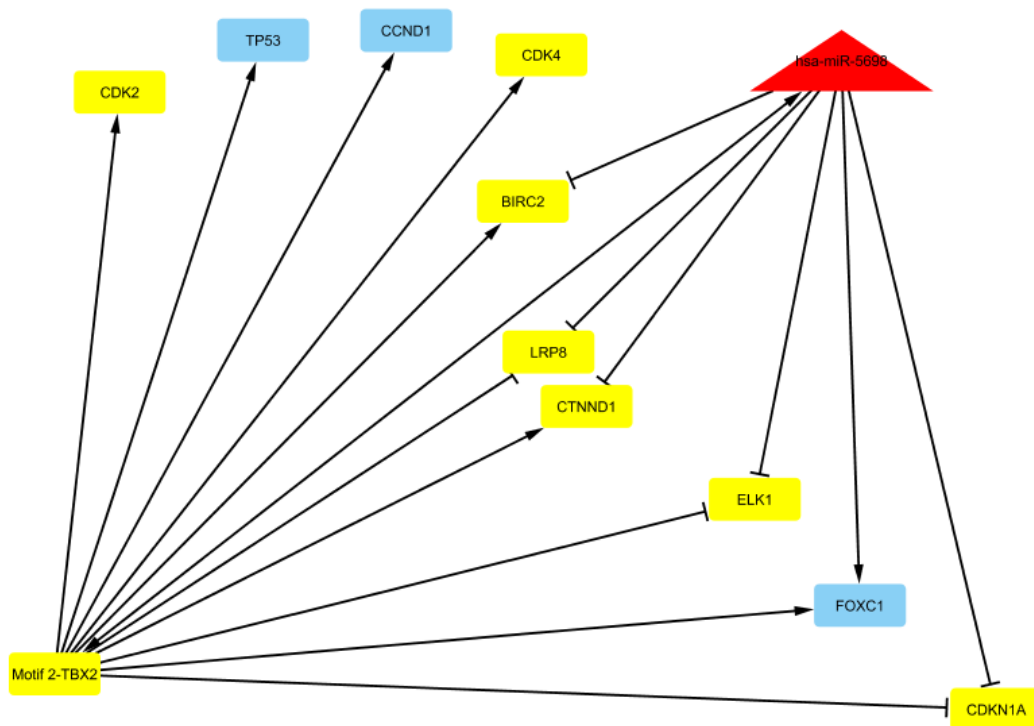


Figure 3.14: Integrative microRNA target genes and the DNA sequence motif 2 transcription factor TBX2 regulatory networks in the prostate metastasis cell line PC3. The Arrows indicate activation of the target by the transcription factor and/or the microRNA. The blunt lines indicate repression of the target by the transcription factor and/or the microRNA. Both the transcription factor and the microRNA activate each other. The genes with the blue coloured nodes are up-regulated in the PC3 cell line, whilst the ones in yellow are down-regulated.

Motif-2 was linked to the transcription factor TBX2. Abnormal expression of this TF has been linked to cell proliferation and invasion in several malignancies including lung cancer (Khalil *et al.*, 2017), ovarian cancer (Tasaka *et al.*, 2017) and nasopharyngeal cancer (Yan *et al.*, 2017). Perhaps this could explain why this TF inhibits CDKN1A in the LNCaP cell line (Figure

3.14) putative regulatory network. The TF activates the FOXC1 gene, perhaps in this environment, the gene takes up its pro-metastatic role.

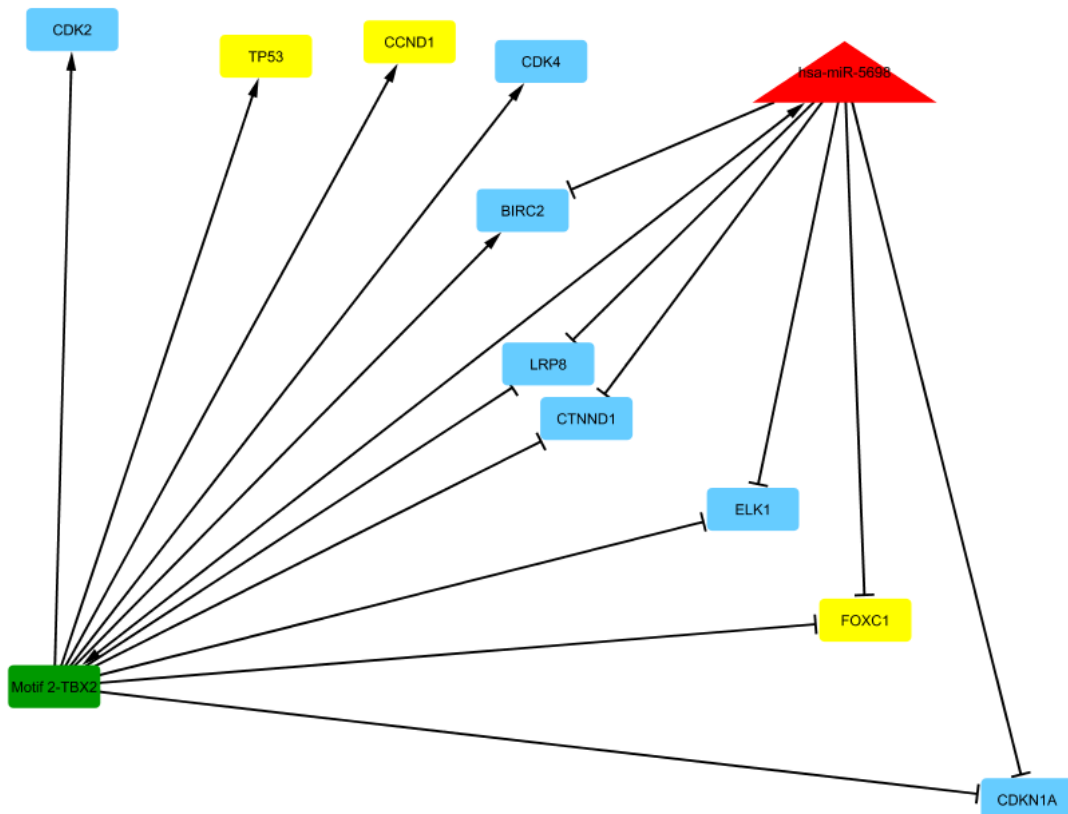


Figure 3.15: Integrative microRNA target genes and the DNA sequence motif 2 transcription factor TBX2 regulatory networks in the prostate adenocarcinoma cell line LNCaP. The Arrows indicate activation of the target by the transcription factor and/or the microRNA. The blunt lines indicate repression of the target by the transcription factor and/or the microRNA. Both the transcription factor and the microRNA activate each other. The genes with the blue coloured nodes are up-regulated in the LNCaP cell line, whilst the ones in yellow are down-regulated.

The putative regulatory network in Figure 3.15, both genes FOXC1 and CDKN1A are inhibited by the TF and the microRNA. This network highlights the complexity of microRNA, TF gene regulation. The qPCR analysis showed that FOC1 is up-regulated in PC3, whilst CDKN1A is down-regulated (Figure 5.1). However, this is not reflected in the network above. Thus, there is a need to perform validation studies via molecular techniques such as microRNA mimic studies, ChIP assays and SELEX assays.

3.3.7 Conclusion

This chapter aimed to identify novel transcription factor binding sites in the promoters of hsa-miR-5698 target genes and investigate their biological significance in the progression of PCa. This was done by first identifying genes co-expressed with the target genes using the databases STRING and IntAct databases, five genes were found to be co-expressed with the microRNA target genes, namely, TP53, CDK4, CDK2, PCNA and CCND1. A set 11 genes including the microRNA targets was the working list. The identification of the cis-regulatory module in the promoters of the working list was accomplished with the MEME Suite tool. Three motifs were identified and their sequences were searched against a database of known transcription factors in the JASPAR database. One top match for each motif was identified namely FOXD3, TBX2 and ZNF263 corresponding to motif 1, 2 and 3 respectively. The expression profiles of the transcription factors was mined using protein expression databases. A correlation analysis was performed to determine the regulatory relationships between the TF,

microRNA and genes using the literature mining data (TF expression profiles), data from Chapter 5 (gene expression profiles) and data from Lombe 2015 (hsa-miR-5698 expression profiles). Two FFLs involved in PCa progression from adenocarcinoma to bone metastasis were identified. A significant feature in the putative regulatory network was the repression of FOXC1 by both TF and microRNA in LNCaP (blunt lines Figure 3.14), but its activation in PC3 (Figure 3.15). Thus, these networks are good candidates for molecular studies.

Transcription factor binding motifs are essential to gene regulatory networks as checkpoints for gene transcription as well as the expression of genes at the right place and time within the cell. In the early 2000's, upon the completion of the sequencing of the human genome, the use of algorithms to predict which transcription factors would bind to particular regions of gene promoters became popular, as it was a cost effective method compared to the wet bench applications that were being employed. The use of algorithmic methods for DNA motif discovery has gradually improved with the validation of many of the binding sites discovered. However, it is not without its drawbacks. The hunt for a common motifs maybe quite complex as the motif sizes can be as small as eight base pairs in length and occur thousands of bases upstream of an unknown subset of genes of interest (D'Haeseleer, 2006). Thus, the use of computational methods to discover motifs should be undertaken with caution.

Thus, several questions can be posed as to the robustness of the methodology used in this study. One such question would be the use of only one predictive algorithm. Given the false positives and negatives of motif discovery tools, it would be wise to use more than one algorithm to

predict a motif. However, the MEME algorithm accounts for this by counting the number of occurrences of all the motifs in a target sequence and calculates the ones most over-represented. Thus, a number of similar over-represented sequences are combined into a more descriptive motif. The algorithm was also run several times and the best scoring motifs of all the iterations were chosen. Additionally, the building of the best motifs for the target sequence was optimized by using a wide range of motif widths and number of sites. Even with the application of these stringent guidelines for the determination of the most probabilistic motif, the results obtained need molecular validation. The current method used is the ChIP-seq assay, a Chromatin immunoprecipitation (ChIP) coupled to high-throughput sequencing.

This study also had drawbacks in the construction of microRNA-TF-gene regulatory networks. This is because the study was limited to the available information available. Although the expression levels of hsa-miR-5698 are available in prostate cancer (LNCaP cell line) from a previous study (Lombe, 2015), information on the transcription factor binding sites and their transcription factors and their effects on the target genes is limited due to their novelty. Additionally, the effect of the transcription factor on the hsa-miR-5698 should also be investigated. Nevertheless, this can be the foundation for a more comprehensive validation of the findings as future work.

3.4 References

1. Abbas, T., Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. *Nature Reviews. Cancer*, 9(6), 400–414. <http://doi.org/10.1038/nrc2657>.
2. Abe, Hideaki, and Neil J Gemmell. "Abundance, Arrangement, and Function of Sequence Motifs in the Chicken Promoters." *BMC Genomics*, 15.1 (2014): 900. PMC.
3. Aerts, S., Thijs, G., Coessens, B., Staes, M., Moreau, Y., and Moor, B.D. 2003. Toucan: Deciphering the cis-regulatory logic of co-regulated genes. *Nucleic Acids Research*. 31(6), 1753–1764.
4. Aken, B.L., Achuthan, P., Akanni, W.M., Amode, R., Bernsdorff, F., Ridwan, A., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P. *et al.*, (2017). Ensembl 2017. *Nucleic Acids Research*. Database issue: D635-D642. PubMed PMID: 27899575; PubMed Central PMCID: PMC5210575. doi:10.1093/nar/gkw1104.
5. Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2010). "Protein Structure and Function". *Essential Cell Biology* (3rd ed.). New York: Garland Science, Taylor and Francis Group, LLC. pp. 120–170.
6. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Noble, W. S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server issue), W202–W208.
7. Bailey, T.L., Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California.

8. Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) Meme: discovering and analysing DNA and protein sequence motifs. *Nucleic Acids Research*, 34, W369–W373.
9. Barash, Y., Elidan, G., Friedman, N., Kaplan, T. (2003). Modeling dependencies in protein-DNA binding sites. Proceedings of the seventh Annual International Conference on Computational Molecular Biology, pp. 28–37.
10. Bhar, A., Haubrock, M., Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S and Wingender, E. (2013). Co-expression and Co-regulation analysis of time-series gene expression data in estrogen-induced breast cancer cell. *Algorithms in Molecular Biology* 8: 9.
11. Bodén, M and Bailey, T.L. (2008). "Associating transcription factor binding site motifs with target Go terms and target genes." *Nucleic Acids Research*, 36, 4108-4117.
12. Bodén, M. and Bailey, T.L. (2008) Associating transcription factor binding site motifs with target go terms and target genes. *Nucleic Acids Research*, 36, 4108–4117.
13. Boeva, B. (2016). Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Frontiers in Genetics*,7:24.
14. Bolton, K.A., Ross, J.P., Grice, D.M., Bowden, N.A., Holliday, E.G., Avery-Kiejda, K.A., Scott, R.J. (2013). STaRRRT: a table of short tandem repeats in regulatory regions of the human genome. *BMC Genomics*, 14:795.
15. Bommer, G.T., Gerin, I., Feng, Y. (2007). p53-mediated activation of miRNA34 candidate tumour-suppressor genes, *Current Biology*, 17 (pg. 1298-307).

16. Buske, F. A., Bodén, M., Bauer, D.C., Bailey, T.L. (2010). Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics*, 26(7), 860-866.
17. Casimiro, A. C., Vinga, S., Freitas, A. T., & Oliveira, A. L. (2008). An analysis of the positional distribution of DNA motifs in promoter regions and its biological relevance. *BMC Bioinformatics*, 9, 89. <http://doi.org/10.1186/1471-2105-9-89>
18. Chohan, T.A., Qian, h., Pan, y., Chen, J.Z. (2015). Cyclin-dependent kinase-2 as a target for cancer therapy: progress in the development of CDK2 inhibitors as anti-cancer agents. *Current Medical Chemistry*, 22 pp. 237-263.
19. Chuanchang, L., Qiangqiang, G., Jiakuan, L., Qingsong, Z., Wang, C., Cui, K., Chen, Z. (2017). Effects of miR-1236-3p and miR-370-5p on activation of p21 in various tumors and its inhibition on the growth of lung cancer cells. *Tumour Biology*; 39:6. <https://doi.org/10.1177/1010428317710824>.
20. Chung, J.H., Bunz, F. (2010). Cdk2 Is Required for p53-Independent G2/M Checkpoint Control. *PLoS Genetic* 6(2). <https://doi.org/10.1371/journal.pgen.1000863>.
21. D'haeseleer, P., (2006). What are DNA sequence motifs? *Nature Biotechnology*. 24 (4), 423-5.
22. Du, W., Fang, Q., Chen, Y., Teng, J., Xiao, Y., Xie, P., Wang, J. (2017). Effect of silencing the T-Box transcription factor TBX2 in prostate cancer PC3 and LNCaP cells. *Molecular Medicine Reports*; 16, 6050-6058. <https://doi.org/10.3892/mmr.2017.7361>.
23. Eastham, J.A., Hall, S.J., Sehgal, I., Wang, J., Timme, T.L., Yang, G., Connell-Crowley, L., Elledge, S.J., Zhang, W.W., Harper, J.W., et al. (1995). In vivo gene therapy with p53 or p21 adenovirus for prostate cancer. *Cancer Research*. 55:5151–5155.

24. Elkon, R., Linhart, C., Sharan, R., Shamir, R., and Shiloh, Y. 2003. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Research*. 13, 773–780.
25. Fagerberg, L., Hallström, B.M., Oksvold P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg E., Szigartyo, C.A., Skogs, M., Takanen, J.O., Berling H, Tegel H, Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular Cell Proteomics*. 13(2):397-406. doi: 10.1074/mcp.M113.035600.
26. Flores, O., Wang, Z., Knudsen, K.E., Burnstein, K.L. (2010). Nuclear targeting of cyclin-dependent kinase 2 reveals essential roles of cyclin-dependent kinase 2 localization and cyclin E in vitamin D-mediated growth inhibition. *Endocrinology*, 151, pp. 896-908, 10.1210/en.2009-1116.
27. 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.
28. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biology*, 8(2):R24, 2007.
29. He, G.Y., Hu, J.L., Zhou, L., Zhu, X.H., Xin, S.N., Zhang, D., Lu, G.F., Liao, W.T., Ding, Y.Q., Liang, L. (2016). The FOXD3/miR-214/MED19 axis suppresses tumour growth and metastasis in human colorectal cancer. *British Journal of Cancer*, 115, 1367–1378 | doi: 10.1038/bjc.2016.362.

30. Hebert, D.N., Molinari, M. (2007). "In and out of the ER: protein folding, quality control, degradation, and related human diseases". *Physiological Reviews*. 87 (4): 1377–408. doi:10.1152/physrev.00050.2006. PMID 17928587.
31. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Apweiler, R. (2004). IntAct: an open source molecular interaction database. *Nucleic Acids Research*, 32(Database issue), D452–D455. <http://doi.org/10.1093/nar/gkh052>
32. Hernandez-Garcia, C.M., Finer, J.J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Science*. Volumes 217–218, Pages 109-119.
33. Jain, A., Raina, K., Agarwal, R. (2013). Deletion of p21/Cdkn1a confers protective effect against prostate tumorigenesis in transgenic adenocarcinoma of the mouse prostate model, *Cell Cycle*, 12:10, 1598-1604, DOI: 10.4161/cc.24741.
34. Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30.
35. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44, D457–D462. doi: 10.1093/nar/gkv1070.
36. Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Hermjakob, H. (2007). IntAct—open source resource for molecular interaction data. *Nucleic Acids Research*, 35(Database issue), D561–D565. <http://doi.org/10.1093/nar/gkl958>.
37. He, G. et al. (2015) 'Hypermethylation of FOXD3 suppresses cell proliferation, invasion and metastasis in hepatocellular carcinoma', *Experimental and Molecular Pathology*, 99(2), pp. 374–382. doi: 10.1016/j.yexmp.2015.06.017.

38. Li, K. et al. (2017) 'FOXD3 is a tumor suppressor of colon cancer by inhibiting EGFR-Ras-Raf-MEK-ERK signal pathway', *Oncotarget*, 8(3), pp. 5048–5056. doi: 10.18632/oncotarget.13790.
39. Tasaka, R., Fukuda, T., Shimomura, S., Inoue, Y., Wada, T., Kawanishi, M., Yasui, T., Sumi, T. (2017). TBX2 expression is associated with platinum- sensitivity of ovarian serous carcinoma. *Oncology Letters*, 15: 3085-3090.
40. Lv, Y., Si, M., Chen, N., Li, Y., Ma, X., Yang, H., Zhang, L., Zhu, H., Xu, G. Y., Wu, G. P., ... Cao, C. (2017). TBX2 over-expression promotes nasopharyngeal cancer cell proliferation and invasion. *Oncotarget*, 8 (32), 52699-52707. doi:10.18632/oncotarget.17084.
41. Khalil, A. A., Sivakumar, S., Lucas, F., McDowell, T., Lang, W., Tabata, K., Fujimoto, J., Yatabe, Y., Spira, A., Scheet, P., Nemer, G., Kadara, H. (2017). TBX2 subfamily suppression in lung cancer pathogenesis: a high-potential marker for early detection. *Oncotarget*, 8(40), 68230-68241. doi:10.18632/oncotarget.19938.
42. Sarkar, S. et al. (2017) 'FOXD3 Regulates CSC Marker, DCLK1-S, and Invasive Potential: Prognostic Implications in Colon Cancer', *Molecular Cancer Research*, 15(12), pp. 1678–1691. doi: 10.1158/1541-7786.MCR-17-0287.
43. Yan, J.-H. et al. (2015) 'FOXD3 suppresses tumor growth and angiogenesis in non-small cell lung cancer', *Biochemical and Biophysical Research Communications*, 466(1), pp. 111–116. doi: 10.1016/j.bbrc.2015.08.116.
44. Yin, H. et al. (2017) 'FOXD3 regulates anaplastic thyroid cancer progression', *Oncotarget*, 8(20), pp. 33644–33651. doi: 10.18632/oncotarget.16853.
45. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Hermjakob, H. (2012). The IntAct molecular interaction database in 2012.

- Nucleic Acids Research*, 40(Database issue), D841–D846.
<http://doi.org/10.1093/nar/gkr1088>
46. Knoll, S., Emmrich S., Pützer, B.M. (2013) The E2F1-miRNA Cancer Progression Network. In: Schmitz U., Wolkenhauer O., Vera J. (eds) *MicroRNA Cancer Regulation. Advances in Experimental Medicine and Biology*, Vol 774. Springer, Dordrecht.
47. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, 172(4), 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>.
48. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L, Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A. (2003) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298 (5594):799-804.
49. Li, J., Di, C., Mattox, A. K., Wu, L. and Adamson, D. C. (2010). The future role of personalized medicine in the treatment of glioblastoma multiforme. *Pharmacogenomics and personalized medicine*, 3, 111.
50. Li, K., Guo, Q., Yang, J., Chen, H., Hu, K., Zhao, J., Li, L. (2017). FOXD3 is a tumor suppressor of colon cancer by inhibiting EGFR-Ras-Raf-MEK-ERK signal pathway. *Oncotarget*, 8(3), 5048–5056.
<http://doi.org/10.18632/oncotarget.13790>.
51. Li, S., Wang, C., Yu, X., Wu, H., Hu, J., Wang, S., Ye, Z. (2017). miR-3619-5p inhibits prostate cancer cell growth by activating CDKN1A expression. *Oncology Reports*; 37:1, pp 241-248.
<https://doi.org/10.3892/or.2016.5250>.

52. Liang, J., Zubovitz, J., Petrocelli, T., *et al.* (2002). PKB/Akt phosphorylates p27, impairs nuclear import of p27 and opposes p27-mediated G1 arrest. *Nature Medicine*; 8: 1153–1160.
53. 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.
54. Lis, M., Walther, D. (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics*, 17, 185. <http://doi.org/10.1186/s12864-016-2549-x>.
55. Lombe, C. (2015). Identification of microRNAs as a class of biomarkers for the early diagnosis of prostate cancer: an *in silico* and molecular approach, MSc thesis, University of the Western Cape, Cape Town.
56. Maclsaac, K.D., Fraenke, E. (2006). Practical Strategies for Discovering Regulatory DNA Sequence Motifs. *PLOS Computational Biology*. <https://doi.org/10.1371/journal.pcbi.0020036>
57. Mariño-Ramírez, L., Tharakaraman, K., Spouge, J.L., and Landsman, D. (2009). Promoter Analysis: Gene Regulatory Motif Identification with A-GLAM. *Methods Molecular Biology* ; 537: 263–276.
58. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., *et al.* (2015). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 44: D110-D115.
59. Meier, S., Tzfadia, O., Vallabhaneni, R., Gehring, C., Wurtzel, E.T. (2011). A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Systems Biology* 19; 5:77.
60. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I. (2005). C-Myc-regulated microRNAs modulate E2F1 expression, *Nature*: 435 (pg. 839-43).

61. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: New, fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research*. (23) 4878–4884.
62. Sadeghi, M., Bijan Ranjbar, Mohamad Reza Ganjalikhany, Faiz M. Khan, Ulf Schmitz, Olaf Wolkenhauer, Shailendra K. Gupta. (2016). MicroRNA and Transcription Factor Gene Regulatory Network Analysis Reveals Key Regulatory Elements Associated with Prostate Cancer Progression. *PLoS ONE* 11(12): <https://doi.org/10.1371/journal.pone.0168760>.
63. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research* 1: 32 (Database issue):D91-4.
64. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M. (2011). "Global quantification of mammalian gene expression control". *Nature*. 473 (7347): 337–42 doi:10.1038/nature10098.
65. Shalgi, R., Lieber, D., Oren M. (2007). Global and local architecture of the mammalian microRNA-transcription factor regulatory network, *PLoS Computational Biology*, 2007, vol. 3 pg. 131.
66. Sharan, R. (2007). *Analysis of Biological Networks: Transcriptional Networks - Promoter Sequence Analysis*, lecture notes, Tel Aviv University, delivered January 4, 2007. Retrieved 25 May 2018.
67. Sunwoo, H.H., Suresh, M.R. (2013). Cancer markers. In Wild, D, ed. *The Immunoassay Handbook. Theory and Applications of Ligand Binding, ELISA and Related Techniques* (4th Edition). New York: Elsevier pgs. 989-1013.
68. Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T, Roth, A., Bork, P., Jensen, L.J., von Mering, C.

- (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*; 45:D362-68.
69. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M. (1999). Systematic determination of genetic network architecture. *Nature Genetics* 22 pp. 281-285
70. Tomilin, N.V (2008). "Regulation of mammalian gene expression by retroelements and non-coding tandem repeats". *BioEssays*, 30 (4): 338–48. doi:10.1002/bies.20741. PMID 18348251.
71. Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., Favorov, A.V., Frith, M.C., Fu, Y., Kent, W.J. et al. (2005) Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites. *Nature Biotechnology*. 23, 137–147.
72. Tsang J, Zhu J, van Oudenaarden A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals, *Molecular Cell*, 2007, vol. 26 (pg. 753-67).
73. Viglietto G, Motti ML, Bruni P, et al. (2002) Cytoplasmic relocalization and inhibition of the cyclin-dependent kinase inhibitor p27kip1 by PKB/Akt-mediated phosphorylation in breast cancer. *Nature Medicine*; 8: 1136–1144.
74. Vences, M.D., Legendre, M., Caldara, M., Hagihara, M., Verstrepen, K.J. (2009). Unstable tandem repeats in promoters confer transcriptional evolvability. *Science*, 324(5931):1213-6.
75. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Research* 1: 31(1):258-61.
76. von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Krüger, B., Snel, B., Bork, P. (2007). STRING 7- Recent developments in the integration

and prediction of protein interactions. *Nucleic Acids Research*. 2007 Jan;35 (Database issue):D358-62. Epub 2006 Nov 10.

77. von Mering, C., Lars J. Jensen, Berend Snel,¹ Sean D. Hooper, Markus Krupp, Mathilde Foglierini, Nelly Jouffre, Martijn A. Huynen,¹ and Peer Bork (2005). STRING: known and predicted protein–protein associations, integrated and transferred across organisms.
78. Wang, C., Huang, Y., Dai, W. (2016). Tumor suppression function of FoxD3 in lung cancer. *Irish Journal of Medical Science*, 185:547–553. DOI 10.1007/s11845-015-1297-2.
79. Wang, Y., Fisher, J.C., Mathew, R., Ou, L., Otieno, S., Sublet, J., *et al.* (2011). Intrinsic disorder mediates the diverse regulatory functions of the CDK inhibitor p21. *Nature Chemical Biology*; 7 (4) (2011), pp. 214-221.
80. Warfel, N.A and El-Deiry, W.S (2013). P21WAF1 and tumorigenesis: 20 years after. *Current Opinion in Oncology*: 25:52–58.
81. Wittkopp, P.J., Kalay, G. (2011). "Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence". *Nature Reviews. Genetics*. 13 (1): 59–69.
82. Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P, Pizer, M., Rockman, M.V., Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology Evolution*. 20(9):1377-419.
83. Xifeng Yin., Jun Yu., Yang Zhou., Chengyue Wang., Zhimin Jiao., Zhounan Qian., Hao., Sun., Binghai Chen. (2018). Identification of CDK2 as a novel target in treatment of prostate cancer. *Future Oncology*, 14(8):709-718.
84. Yan, J.H., Zhao, C.H., Ding, L.B., Zhou, X (2015). FOXD3 suppresses tumor growth and angiogenesis in non-small cell lung cancer. *Biochemical and Biophysical Research Communication*. 466, 111-116.

85. Yang, D., Sun, Y., Hu, L., Zheng, H., Ji, P., Pecot, C.V. (2013) Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer Cell* 23: 186–199.
86. Yee, K.Y ., Medvedovic, M., Bumgarner, R.E. (2004). From co-expression to co-regulation: how many microarray experiments do we need? *Genome Biology* 5:R48.
87. Yilmaz, A., Grotewold, E. (2010). Components and Mechanisms of Regulation of Gene Expression. In: Ladunga I. (eds) *Computational Biology of Transcription Factor Binding. Methods in Molecular Biology* (Methods and Protocols), 674. Humana Press, Totowa, NJ.
88. Yu, H., Luscombe, N., Qian, J., and Gerstein, M. (2003). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends in Genetics* 19(8):422-7.
89. Yuan, L., Chen, L., Qian, K., Qian, G., Wu, C., Wang, X., Xiao, Y. (2017). Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genomics Data*. Volume 14. Pages 132-140.
90. Zhang, H-M., Shuzhen Kuang, Xushen Xiong, Tianliuyun Gao, Chenglin Liu, An-Yuan Guo. (2015). Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases, *Briefings in Bioinformatics*, 16;1, Pages 45–58, <https://doi.org/10.1093/bib/bbt085>.

Chapter 4

An *in silico* prognostic analysis of hsa-miR-5698 target genes in prostate cancer (PCa)

4.1 Background

A prognostic biomarker indicates an increase or decrease in the likelihood of a future clinical event, disease occurrence and/or its progression in a particular population (Roberts *et al.*, 2001). In order to measure a prognostic biomarker, a starting point or baseline needs to be defined. This means that monitoring of a particular marker can be done in individuals at risk for the disease, after diagnosis with no treatment, after treatment has commenced as well as between various types of treatment (FDA-NIH Biomarker Working Group, 2016; Tang *et al.*, 2017).

There are a number of markers used to indicate prognosis of a disease in clinical oncology, these include tumour size, number of lymph nodes positive for tumour cells, and presence of metastasis (Nalejska *et al.*, 2014). In addition to these clinical pathologic characteristics, there has been an increase in the use of molecular signatures measured on tumours, as indicators. In prostate cancer these include, elevated Prostate Specific Antigen (PSA) level at diagnosis, more advanced clinical and pathological tumour staging (Buhmeida *et al.*, 2006; Martin *et al.*, 2011) and higher Gleason score (D'Amico *et al.*, 1998; D'Amico *et al.*, 2003; Terada *et al.*, 2017). There have been a number of nomograms developed with both

clinical pathologic characteristics and molecular signatures to predict PCa outcome (Ross *et al.*, 2001; Shariat *et al.*, 2009). One such nomogram developed by Kattan and co-workers (1995) uses the aforementioned characteristics and signature to predict biochemical recurrence (PSA) after a prostatectomy.

The clinical course of prostate cancer is long, particularly when the five to ten year lead-time added by PSA screening is taken into account. Many studies (Pound *et al.*, 1999; Bianco *et al.*, 2005; Terada *et al.*, 2017) have used a rise in PSA following local therapy to assess the risk of disease progression. However, this is not a perfect surrogate for prostate cancer mortality, as the vast majority of these patients do not succumb to their disease.

4.1.1 Prognostic determinants in Prostate cancer (PCa)

After the introduction of Prostate Specific Antigen (PSA) as a biomarker for the evaluation of the clinical course of prostate cancer, the search for prostate cancer biomarkers has shifted towards the identification of gene or protein expression levels capable of predicting the prognosis of the disease (Tahara *et al.*, 2017). These expression signatures are useful in the evaluation of the best course of treatment and therapy after surgical resection (Fizazi *et al.*, 2011). The identification of genes, whose expression signatures and protein products could serve as biomarkers for prostate cancer outcomes, would be a useful addition to the current methods used in the management of the disease. The following review outlines the

current methods used in the management of PCa. It examines their usefulness and shortcomings in the management of the disease.

4.1.2 Clinical factors

4.1.2.1 Serum Markers - Prostate Specific Antigen (PSA)

Regular prostate cancer screening programs such as the digital rectal exam and the monitoring of the PSA biomarkers provide a means of early detection and bring about better disease prognosis, which also contributes to improved outcomes. However, PSA is organ-specific, but not cancer-specific and it is not able to differentiate between indolent and aggressive forms of prostate cancer (Terada *et al.*, 2017; Velonas *et al.*, 2013). Thus, not all low or high PSA levels are indicative of prostate cancer. Studies also show that many men may harbour aggressive prostate cancer despite having low initial levels of serum PSA (Caram *et al.*, 2016; Slatkoff *et al.*, 2011; Thompson *et al.*, 2004).

There have been a number of potential serum markers other than PSA that have been investigated for their role in providing prognostic information in PCa. One such marker is the transforming growth factor beta gene (TGF- β) which has a significant role in the tumour microenvironment (Tu *et al.*, 2003). This gene secretes a protein that performs many cellular functions, including the control of cell growth, cell proliferation, cell differentiation and apoptosis in early tumourigenesis (Bello-DeOcampo and Tindall 2003; Dos Reis *et al.*, 2011; Cao and Kyprianou, 2015). However, in later stages of

tumour progression, it acts as a tumour promoter. Over-expression of TGF- β aids in tumorigenesis by stimulating angiogenesis and suppression of the immune system causing the cancer cells to become resistant to TGF- β induced inhibition and apoptosis (Bello-DeOcampo and Tindall 2003; Dos Reis *et al.*, 2011). Thus, high expressions of TGF- β in later stages of prostate cancer is associated with poor clinical outcome. However, the gene performs the same in many cancers including but not limited to breast cancer (de Kruijf *et al.*, 2013; Zarzynska, 2014) and colorectal cancer (Chun *et al.*, 2017; Jung *et al.*, 2017).

4.1.2.2. Gleason score

As discussed earlier (section 3.1), a number of markers including clinical pathologic characteristics and molecular signatures measured on tumours are used to indicate prognosis of PCa. One of the clinical factors used in prognosis is the Gleason grade. The Gleason grade characterizes prostate tumour architecture and morphology via the assignment of a primary and secondary score with a summation of the total score within a range of 2 to 10 (Gleason and Mellinger, 1974). Men diagnosed with Gleason grade 7 or higher tumours are at increased risk of extra-prostatic extension, increased risk of recurrence after initial therapy, and more likely to die of their disease (Buhmeida *et al.*, 2006; Martin *et al.*, 2011). In contrast, men diagnosed with well-differentiated Gleason 6 disease score are at very low risk of cancer-specific death. The distribution of Gleason grades has shifted over time (Albertsen *et al.*, 2005; Stark *et al.*, 2009), and in the era of PSA

screening, most men are now diagnosed with Gleason 6 or 7 tumours. As such, the accurate discrimination of prognosis among men with prostate cancer within this narrow range of Gleason scores is challenging.

4.1.3. Molecular factors

For many years, clinical factors have proven useful for risk identification and guidance of treatment decisions in prostate cancer. However, they are prone to significant clinical heterogeneity and as such; the exploration of molecular and genetic factors has been spearheaded to improve risk prediction (Buhmeida *et al.*, 2006; Martin *et al.*, 2011). An ideal prognostic marker should be measurable at diagnosis, have both high sensitivity and specificity for distinguishing indolent and lethal PCa and be cost effective (Ludwig and Weinstein, 2005).

4.1.3.1 Tumours markers

There have been a number of tumour related markers that have been studied as potential prognostic markers in prostate cancer. Early studies were based on small series of clinical cohorts. However, they did not have sufficient statistical power to detect meaningful associations (Martin *et al.*,

2011). Recently, protein expression of immuno-histochemical biomarkers has been widely utilized for prognostic value.

(i) Tumour markers involved in evasion of apoptosis

One of the hallmarks of cancer is resisting cell death. The process of apoptosis is well regulated and mediated by a number of factors many of which have been explored as prognostic factors in prostate cancer. One such example is the protein P53, coded for by the gene TP53. This gene is regularly mutated in many cancers, however, in prostate cancer less frequently so (Kan *et al.*, 2010). The potential prognostic utility of P53 lies in its localization. When sequestered in the nucleus, it indicates a stabilized mutation of the protein and has been associated with poor prostate cancer prognosis (O'Brate and Giannakakou, 2003; Maki, 2010).

(ii) Self-sufficiency in growth signals

Tumour cells develop the ability to promote growth in the absence of signalling systems that are normally regulated. There are a number of signalling pathways that are inter-related and involved in promotion of growth. These include the androgen signalling pathway and the epidermal growth factor receptor (EGFR). Several studies (Di Lorenzo *et al.*, 2002;

Baselga *et al.*, 2005; Schlomm *et al.*, 2007) have shown that the EGFR pathway is a prognostic tumour factor. An increase in EGFR staining and copy number is associated with an increased risk of recurrence following prostatectomy. The androgen receptor signalling pathway has great prognostic value (Shukla-Dave *et al.*, 2009). Taplin and co-workers found that mutations in the pathway are common in the castration resistant setting and are additionally associated with response to treatment. Later studies have supported this claim (Taplin *et al.*, 1995; Loneragan and Tindall, 2011; Tan *et al.*, 2015).

(iii) Signature markers of lethal prostate cancer

There are a number of signature markers for prostate cancer prognosis. These have been elucidated in many studies. One such study, conducted by Ding and co-workers found that a loss of the PTEN gene as well as the SMAD4 gene within the prostate resulted in the development of metastatic prostate cancer (Ding *et al.*, 2011). Upon comparing mRNA expression profiles of tumours with PTEN and those without, they found alterations in several pathways including proliferation and invasion as well as metastasis. Additionally, they found that loss of the PTEN and SMAD4 coupled with high expression of CCND1 and SPP1 genes is a signature of poor prognosis in prostate cancer, being associated with biochemical recurrence and lethal prostate cancer. Although the four-marker signature shows better prediction of lethal prostate cancer than the Gleason score and the PSA,

more work is needed to validate the markers in a larger cohort of samples annotated with prostate cancer specific deaths.

4.1.3.2 Urine markers

With the urethra, being in close proximity to the prostate, urine may also be considered a prospective source of clinically useful biomarkers in men with prostate cancer. The prostate cancer antigen 3 (PCA3) is a gene that codes for the PCA3 protein. This protein has been found to be differentially expressed in prostate cancer compared to normal prostate tissue (Bussemakers *et al.*, 1999; Marks and Bostwick, 2008). With an expression which is 60 fold greater in cancerous than benign prostate tissue (Merola *et al.*, 2015). The prostate cancer antigen 3 is shed from the prostate tumour into the urine and its mRNA can be measured in the first void urine after a digital rectal exam (DRE) and a score associated with this measurement (Marks and Bostwick, 2008; Merola *et al.*, 2015). Thus, PCA3 can improve upon the specificity of protein specific antigen (PSA).

Wei and co-workers (2015) performed a multivariate analysis on the association of PCA3 score to currently used indicators of prognosis (Gleason score, PSA and clinical stage). They found that a high PCA3 score in urine was significantly correlated with a high Gleason score as well an advanced clinical stage. Therein lies the ability of PCA3 to act as a prognosis biomarker for PCa. However, the collection of PCA3 is dependent on the tumour cells exfoliation. Thus, a prostate massage is required before sample collection and this is achieved via a digital rectal exam. This is an

invasive approach and may go against some advantages of the ease of urinary sample collection.

Currently, there is a need for sensitive prognostic markers for prostate cancer and based on the review, molecular based signature markers are a promising tool for risk stratification in patients. Molecular based markers might particularly gain wider application if translated to easy-to-use procedures such as immunohistochemistry.

4.1.3.3 The role of bioinformatics in biomarker discovery

High throughput technology platforms in proteomics and genomics have accelerated the development of biomarkers. Furthermore, recent successes of several new agents in PCa treatment, such as immunotherapy, have stimulated the search for predictors of response and resistance. This chapter investigates the potential of hsa-miR-5698 and its target genes to serve as prognostic biomarkers in prostate cancer. The metastatic ability of the genes in prostate cancer will also be investigated. The study was conducted using *in silico* approaches.

The high throughput technology platforms have led to the generation of a large number of datasets. These can be used to carry out various studies including validation, statistical analysis as well as meta-analyses of biomarkers. A number of datasets from online platforms were employed to perform prognostic analysis for this study, these included, ProgGene, SurvExpress, and the Human Cancer Metastasis Database (HCMD).

4.2.1 ProGgene

The database ProGgene is a web application that can be used for studying prognostic implications of mRNA biomarkers in a variety of cancers. The database is compilation of data from public repositories such as GEO, EBI Array Express and The Cancer Genome Atlas. It can be launched from <http://www.compbio.iupui.edu/proggene>. ProGgene has 64 patient series from 18 cancer types in its database providing the most comprehensive resource available for survival analysis to date (Goswami and Nakshatri, 2013). The ProGgene database accepts lists with Official Gene Symbol identifiers and survival measures can be analysed for metastasis, relapse and death. Returned results are based on a Kaplan Meier plot for the risk groups.

4.2.2 SurvExpress®

SurvExpress® available at <http://bioinformatica.mty.itesm.mx/SurvExpress> is a web based tool that provides survival analysis and risk assessment of cancer datasets. The SurvExpress® database comprises 20,000 samples from 20 different cancers curated in 130 datasets (Aguirre-Gamboa *et al.*, 2013). The database is a bioinformatics tool that examines the performance of mRNA biomarkers for survival and prognostic outcomes in various cancers.

The Survexpress® database accepts gene lists in the following identifiers Entrez, Official Gene Symbol, Ensembl, HGNC, MIM, Vega and HPRD. SurvExpress® is the largest and the most versatile free tool for analysing differentially expressed multi-gene biomarkers in human cancers at the same time (Aguirre-Gamboa *et al.*, 2013). Analysis in the database is based on a Kaplan Meier plot for risk groups, clinical information available related to risk group, heat map representation of the gene expression values, a box plot across risk groups, and tables with the summary of the Cox fitting and the prognostic indices (Aguirre-Gamboa *et al.*, 2013).

4.2.3 HCMDB: the human cancer metastasis database

Metastasis is the spread of a cancer from one organ to another without being directly connected with it, and it is the principal cause of cancer-related death (Fokas, 2007). The human cancer metastasis database (HCMDB) is a freely accessible platform that aids in the query of transcriptome data on metastases obtained from different platforms. It is available from <http://hcmdb.i-sanger.com/index>.

The database was created from 620 datasets from the Gene Expression Omnibus (GEO) and the Sequence Read Archive (SRA) containing data on primary tumours and metastases. In addition, clinical data corresponding to metastasis-related expression was retrieved from The Cancer Genome Atlas (TCGA) and included in the database. Currently, the database has gene expression profiles of 29 primary tumour types from 455 experiments with a total of 11,500 samples (Zheng *et al.*, 2018).

4.3 Aims and Objectives

The microRNA hsa-miR-5698 was found to be up-regulated in prostate cancer in a previous study conducted in our lab. This chapter aims to investigate the signature of the microRNA's targets (CDKN1A, CTNND1, ELK1, BIRC2, FOXC1, and LRP8) as molecular prognostic markers for prostate cancer (PCa) using *in silico* methods.

Specific objectives:

1. Evaluate the association between gene targets of hsa-miR-5698 and PCA patient survival, recurrence and metastasis using Kaplan-Meier analysis with the Log-rank test in the SurvExpress and ProgGene databases.
2. Determine the effective prognostic ability of the hsa-miR-5698 targets in terms of metastasis in PCA using the Human Cancer Metastasis Database.

4.4 Methodology

4.4.1. Log-rank survival, recurrence and metastasis analysis of the prognostic value of the target genes via SurvExpress

The online biomarker validation tool SurvExpress was accessed online at <http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp>. The six microRNA target genes' official gene symbols were used as input in the space provided for gene list and prostate selected as the tissue of analysis. Several datasets were used to analyse the genes as survival, recurrence and metastatic markers. The datasets used for survival analysis were the PRAD - TCGA - Prostate adenocarcinoma dataset, the Galsky - Prostate - GSE45705 and the Kollmeyer-Jenkins Prostate GSE10645-GPL5858 each curating 497, 61 and 596 samples respectively.

The datasets used for recurrence analysis were the Lapointe Prostate PNAS, Taylor MSKCC Prostate and the Singh Prostate Nature each curating 29, 140 and 98 samples respectively.

Metastatic analysis was undertaken on the Lapointe Prostate PNAS dataset curating 28 samples (Aguirre-Gamboa *et al.*, 2013). The analysis button was clicked and on the next page, "Survival, metastasis or recurrence was selected and the output downloaded in PDF format.

4.4.2 Log-rank relapse free survival and death analysis of the prognostic value of the target genes via ProGgene

The database ProGgene was launched from

<http://watson.compbio.iupui.edu/chirayu/proggene/database/?url=proggene>. The six microRNA target genes' official symbols were used as input in the space provided for gene list and prostate selected as the tissue of analysis. The databases used for relapse free survival was the GSE70769 with 281 sample and The Cancer Genome Atlas (TCGA) dataset on prostate carcinoma with a data size of 498. The dataset used for relapse free survival was the GSE16560 with a data size of 264 and The Cancer Genome Atlas (TCGA) dataset on prostate carcinoma (size 498). The analysis measure was done separately for relapse and death.

4.4.3 Evaluation of the metastatic prognostic ability of the target genes using the Human Cancer Metastasis Database (HCMD)

The Human Cancer Metastasis Database was launched from the following URL, <http://hcmdb.i-sanger.com>. The official gene symbol of the six target genes were used as input into the space provided for the analysis and the cancer type prostate adenocarcinoma selected. The results returned are also available for download from <http://hcmdb.i-sanger.com/download>.

4.5 Results and discussion

The prognostic value of the hsa-miR-5698 target genes in terms of survival, recurrence and metastasis was examined using the methods outlined in Section 4.4.

4.5.1 Log-rank metastatic analysis of the prognostic value of the target genes in prostate cancer via SurvExpress

The prognostic potential of the genes was evaluated using the Lapointe Prostate PNAS dataset. Only results for four out of the six genes were returned for the event metastasis in the dataset, these were CDKN1A, CTNND1, BIRC2 and FOXC1. The dataset examined two cohorts, the prostate cancer high risk group (red) and the prostate cancer low risk group (green).

With regards to expression of the candidate genes, it can be seen from Figure 4.1 that the genes FOXC1, BIRC2 and CDKN1A are significantly up-regulated in the low risk group when compared to the high risk group as most of the samples expressing these genes lie in the upper quartile. On the other hand, there is no significant difference of CTNND1 expression in both cohorts. Thus, the three genes, CDKN1A, BIRC2 and FOXC1 may serve as good potential markers for metastatic prognosis in prostate cancer.

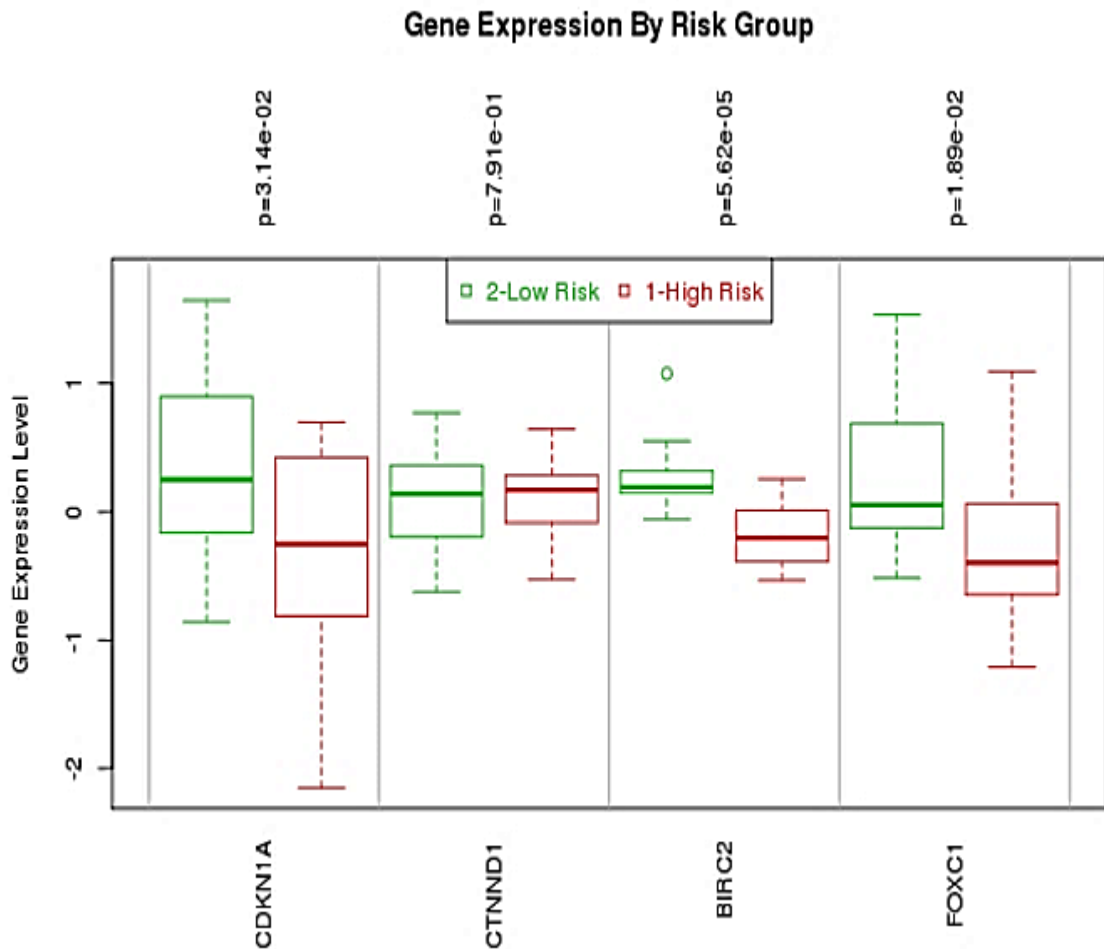


Figure 4.1: Gene expression values for each of the genes in each cohort (High risk red and Low risk green) represented by box and whisker plots. Lapointe Prostate PNAS dataset.

Figure 4.2 shows the Kaplan-Meier plots by risk group, the log-rank test of differences between risk groups, the hazard-ratio estimate, and the concordance indices for the genes in the Lapointe Prostate PNAS dataset. From this figure, it can be seen that more candidates in the low-risk cohort (green) experience metastasis (event) when compared to the candidates in the high-risk cohort (red). However, after the median survival time, more

candidates from the high-risk cohorts who express the markers of interest significantly lower than their low-risk counterparts, experience the event and by the end of the study period, all the high-risk candidates have expired.

An examination of the roles of the markers of interest could help in understanding their behaviour in this study. The gene CDKN1A as mentioned in section 3.3.2 is a Cyclin dependent protein that is an inhibitor of cell cycle progression in the G1/S and G2/M transitions. Up-regulation of this gene inhibits prostate cancer tumourigenesis. Perhaps, this could indicate why there are less candidates experiencing the event in the low-risk group compared to their counterparts in the high-risk cohort, as the former expresses the marker significantly higher than the latter (Figure 4.1).

The gene FOXC1 is a forkhead box C1 transcription factor. It has recently been shown to have profound and critical roles in the progression of several cancer types. This is following its initial identification as a key prognostic indicator of basal-like breast cancer (Han *et al.*, 2015; Jin *et al.*, 2015). Aberrant expression of this gene has been linked to malignancy, proliferation, differentiation, survival and metastasis. (Yang *et al.*, 2017; Elian *et al.*, 2018). Overexpression of FOXC1 has been known to inhibit invasive progression in prostate cancer (Yang *et al.*, 2017). Our data in Figure 4.1 shows that it is highly expressed in the low-risk cohort.

The gene FOXC1 is also involved in cell invasion and metastasis in many cancers including prostate cancer (Kalluri *et al.*, 2009). FOXC1 is consistently elevated by the overexpression of TGF- β 1, snail and twist in the epithelial–mesenchymal transition pathway (EMT) (Batlle *et al.*, 2000; Xu *et al.*, 2012; Huang *et al.*, 2015). Activation of this pathway characterises

cancer metastasis (Kalluri *et al.*, 2009). This dual effect of FOXC1 as both a tumour suppressor and pro-metastatic mediator makes it an interesting target for further analysis.

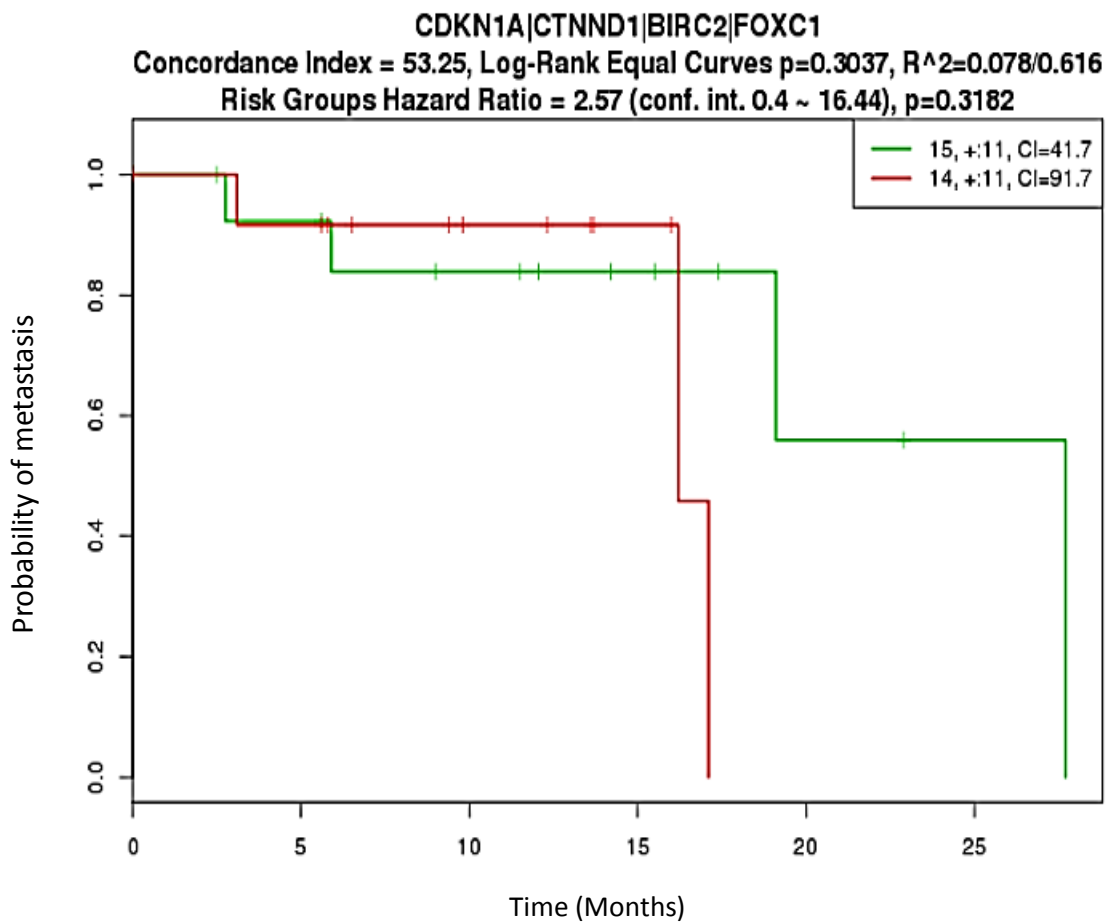


Figure 4.2: Combined Kaplan Meier analysis for four candidate genes for prostate cancer metastasis in the SurvExpress database. The X-axis represents time in months. The Y-axis represents the percentage probability of metastasis in the patients. Lapointe Prostate PNAS database.

The heat map shown in Figure 4.3 ranked the genes based on their prognostic ability. It showed that most of these genes are highly differentially expressed in the low-risk group. This data, coupled with the knowledge of the function of the said genes in prostate cancer may be an indication that their high expression might present a good prognostic outcome in prostate cancer patients with regards to metastasis. However, results from a qPCR analysis (Chapter 5, Figure 5.1) shows that in the metastasis prostate cell line PC3, the genes FOXC1 and CDKN1A are up-regulated and the genes BIRC2 and CTNND1 are down-regulated. On the other hand, a caveat to this result is that cell lines are a snap shot of a single event, unlike tumours which are heterogeneous for PCa.

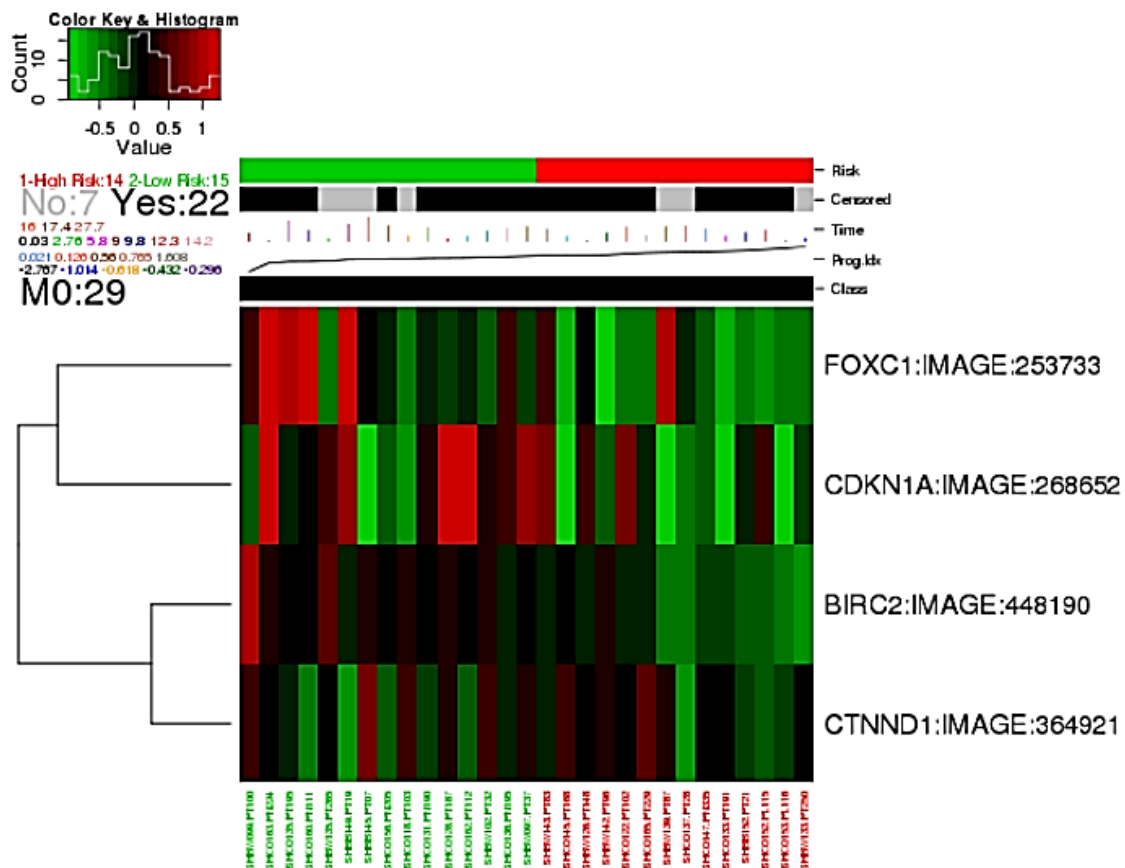


Figure 4.3: Heat map demonstrating expression of each of the candidate genes in each of the 29 samples used in the dataset. According to the colour key, red represents up-regulated genes and green represents down-regulated genes. Black represents unchanged expression. Lapointe Prostate PNAS database.

In patients with metastatic prostate cancer, the median survival rate is approximately 30 months. This is reflected in our Kaplan Meier plot. However, it has been reported that the approximate median survival is dependent upon various factors including treatment with hormone therapy (American Society of Clinical Oncology, 2018). Initial studies showed that

there is substantial inter-patient variation with this therapy as well as other therapies (Glass *et al.*, 2003; Gravis *et al.*, 2015). Thus, there are a number of major prognostic factors that have a major impact on patient outcome, that should be taken into consideration when examining this data, namely, appendicular versus axial disease, age, prostate specific antigen less than 65 versus 65 ng/ml or greater and Gleason score less than 8 versus 8 or greater Glass *et al.*, 2003).

4.5.2 Log-rank relapse free survival (recurrence) analysis of the prognostic value of the target genes in prostate cancer via ProGene.

The prognostic potential with regards to recurrence for the genes was evaluated using the GSE70769 dataset. Results for all of the six genes were returned for the event relapse free survival. The dataset examined two cohorts, the high expression of the biomarker in prostate cancer (red) and the low expression of the biomarker in prostate cancer (green).

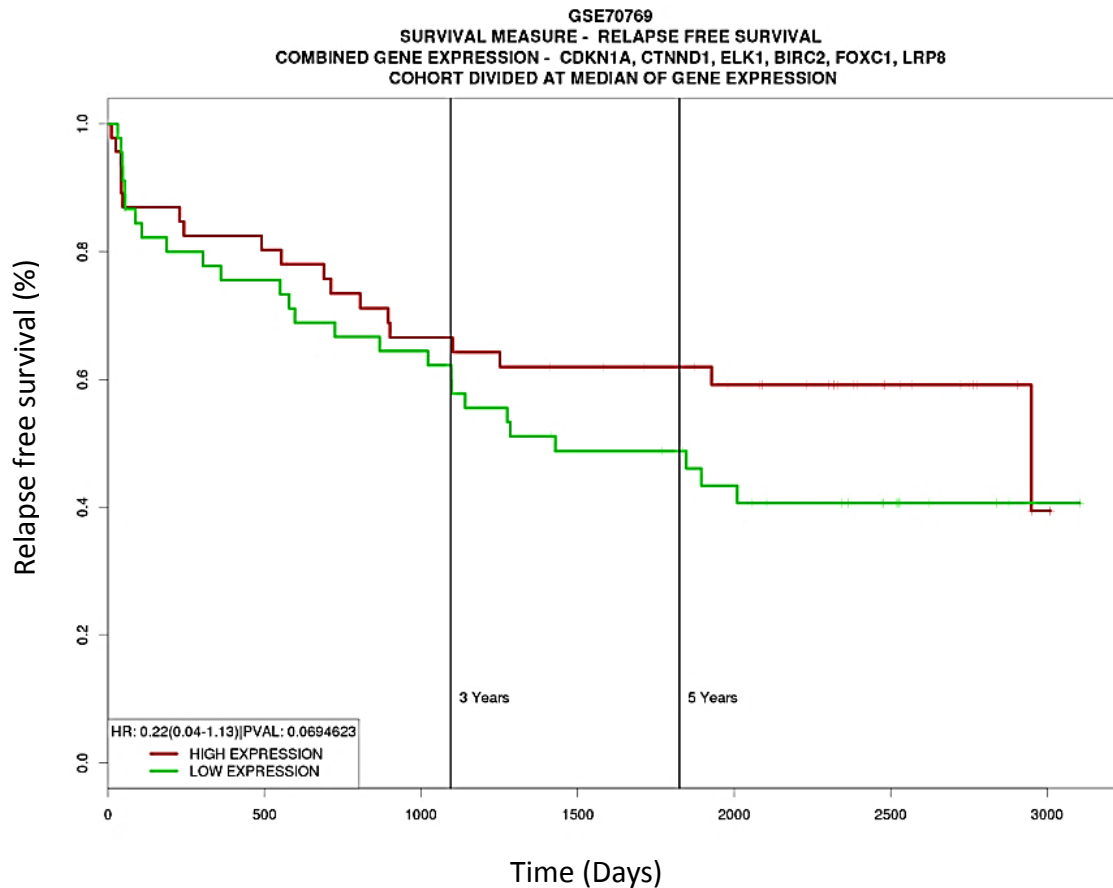


Figure 4.4: Combined Kaplan Meier analysis plot all microRNA target genes for prostate cancer relapse free survival in the ProGgene server. The X-axis represents time in days. The Y-axis represents the percentage probability of relapse free survival of the patients. GSE70769 dataset.

Figure 4.4 shows the Kaplan-Meier plot by biomarker expression groups, which estimate the probability that subjects with a low expression (green) will experience relapse free survival (event) before subjects with a higher expression (red) for the GSE70769 dataset. From this figure, we can see that more candidates in the low biomarker expression cohort experience

the event when compared to the candidates in the high expression cohort. More than half of the patients not expressing the biomarker have had a relapse of the disease.

One of the genes represented in this combined Kaplan Meier plot is ELK1 which is a direct target of hsa-miR-5698. It codes for a transcription factor that is known to be involved in a number of functions including regulation of cell proliferation, cell cycle and apoptosis (Hipskind *et al.*, 1991; Shao *et al.*, 1998; Zhang *et al.*, 2013). In prostate cancer cell lines the androgen receptor (AR), which is a key driver of tumour progression has been suggested to function as a co-activator of ELK1. Silencing ELK1 retards the proliferation of AR- positive cells (Patki *et al.*, 2013). A qPCR analysis of this gene in LNCaP and PC3 cell lines (Chapter 5 Figure 5.1) showed that it is up-regulated in the adenocarcinoma LNCaP cell line but down-regulated in the metastatic PC3 cell line. It can also be seen that low expression of this gene is a good prognosis for relapse free survival (Appendix A, Figure 7.2 (1)).

Another gene that is represented in the combined Kaplan Meier plot and is a direct target of our microRNA of interest is the LRP8 gene. This gene codes for the Lipoprotein Receptor-Related Protein 8. In our qPCR analysis, this gene was found to be up-regulated in both the LNCaP and PC3 cell lines. High expressions for this gene in prostate cancer is a poor prognosis for overall survival (Appendix A, Figure 7.3 (2)).

4.5.3 Log-rank overall survival analysis of the prognostic value of the target genes in prostate cancer via ProGgene.

The prognostic potential with regards to overall survival for the genes was evaluated using the GSE16560 dataset and the PRAD dataset in The Cancer Genome Atlas (TCGA) database available from the ProGgene server platform. Results for all of the six genes were returned for the event overall survival. The dataset examined two cohorts, the high expression of the biomarker in prostate cancer (red) and the low expression of the biomarker in prostate cancer (green) (Figure 4.6).

From Figure 4.5, it can be seen that prostate cancer overall survival is long, with some candidates from the study surviving to over 15 years after diagnosis in both cohorts. The same is seen in Figure 4.6 with data from the TCGA database on the same event. At the median survival time (y-axis 0.5) we see that after three years, approximately more than 80 % of the patients have not experienced the event. After five years, approximately 50 % of the candidates are still alive (Figure 4.5). At the end of the study, members of the cohort with a high expression of our markers of interest have expired in both datasets examined.

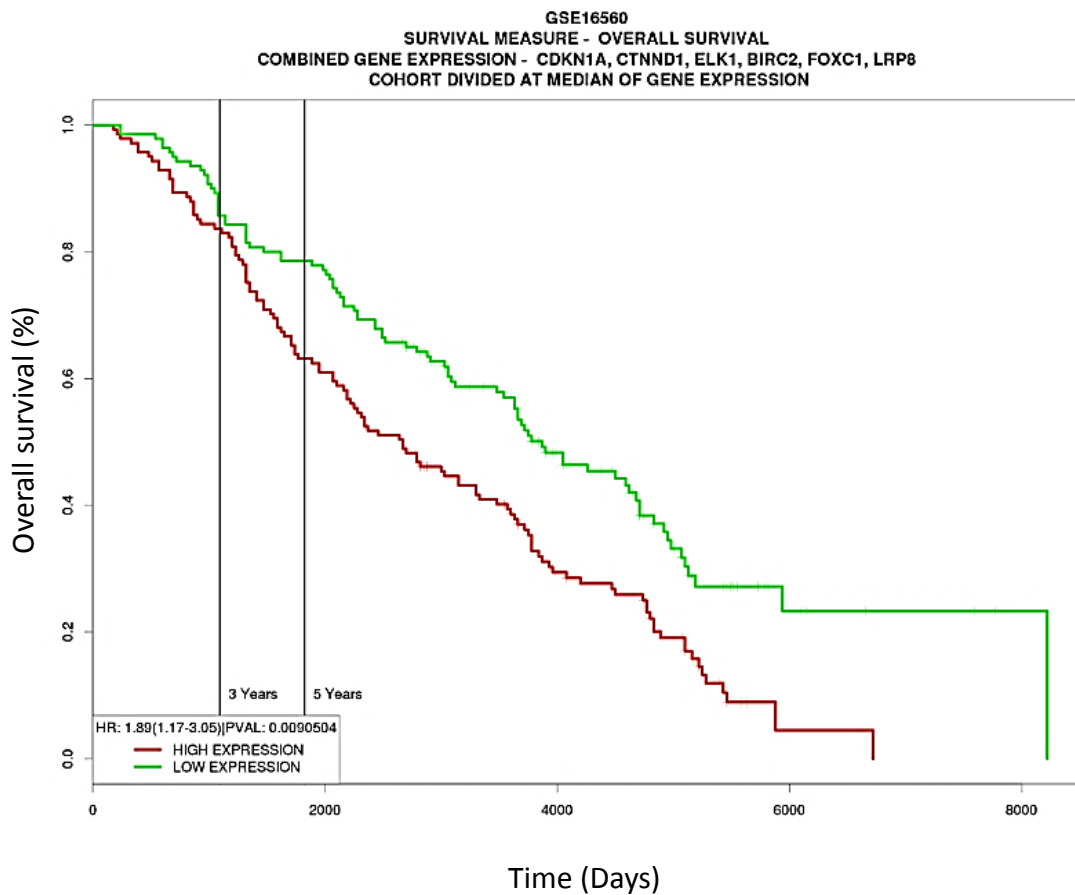


Figure 4.5: Combined Kaplan Meier analysis plot of all microRNA target genes for prostate cancer overall survival in the ProGene server in the GSE16560 dataset.

In Figure 4.6, analysis of the same event in a different dataset shows that at the median survival time, after three years, an estimate of more than 95 % of the candidates in both cohorts have not experienced the event. The same holds for the five-year mark. The relative survival rate for localized or regional prostate cancer is 98 % after the first five years (American Society of Clinical Oncology, 2018).

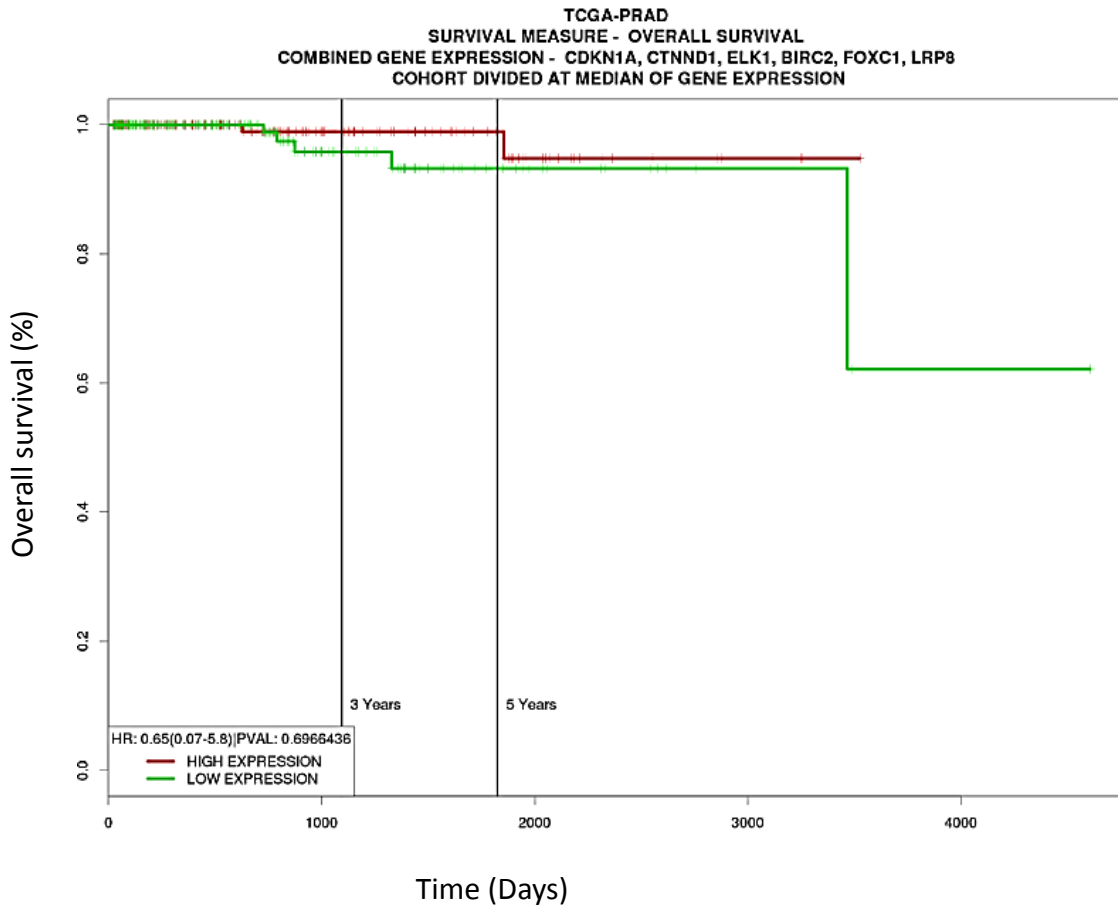


Figure 4.6: Combined Kaplan Meier analysis plot of all microRNA target genes for prostate cancer overall survival in the ProGgene server. The X-axis represents time in days. The Y-axis represents the percentage probability of overall survival of the patients. The Cancer Genome Atlas dataset.

4.5.4 Metastatic prognostic ability of the target genes using the Human Cancer Metastasis Database (HCMD)

The metastatic prognostic ability of the genes was examined in the HCMD. Two datasets were examined for this study, the EXP00437 and the EXP00438

datasets each containing 494 and 542 samples respectively. The EXP00437 dataset was used to examine the primary tumours with metastasis versus primary tumours without metastasis, whilst the EXP00438 dataset was used to examine normal prostate tissue with primary tumour for all the genes. The data is a collection of several microarray experiments normalised in the National Cancer Institute database and curated in The Cancer Genome Atlas. The sites of metastasis explored were bone, non-regional and distant lymph nodes. The results obtained are represented in Figure 4.7.

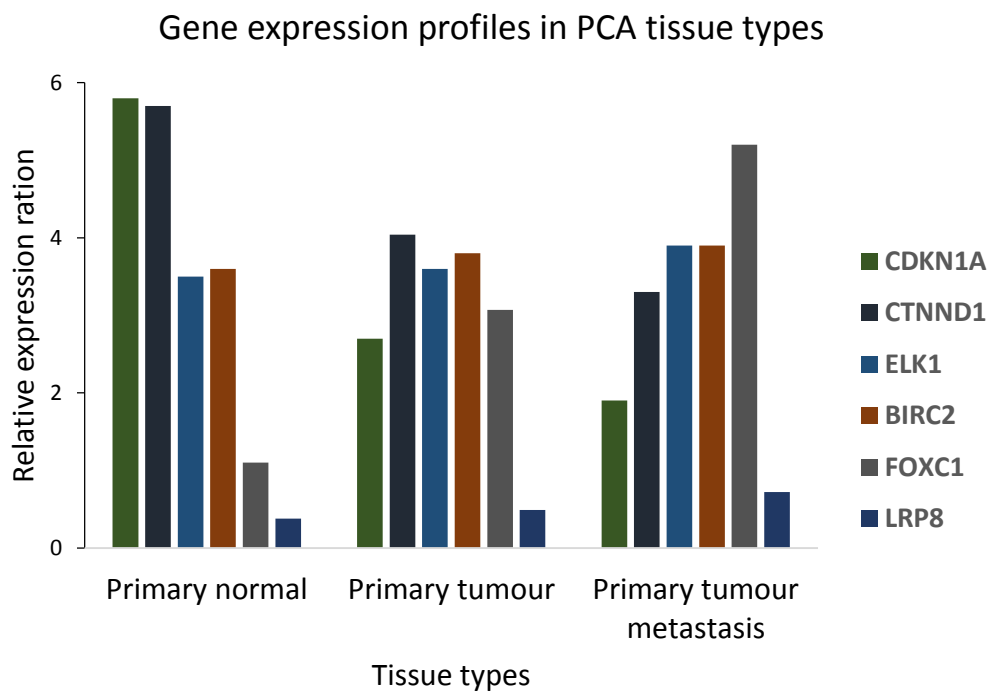


Figure 4.7: Comparison of gene expression profiles in prostate normal tissue, prostate adenocarcinoma primary tumour and metastatic tumour from The Human Cancer Metastasis Database (HCMD).

An examination of the primary tumour metastasis column in Figure 4.7 shows that the genes CDKN1A and CTNND1 are significantly down-regulated in metastasis when compared to normal expression and in the primary tumour. From the survival analysis plots on metastasis in SurvExpress (Figure 4.1 and 4.2), it can be seen that a high expression of these genes in patients with prostate cancer is a good prognosis of metastasis for the disease.

For the genes ELK1 and BIRC2, we see that their expression in primary tumour tissue is higher than in normal tissue and even slightly higher in the metastatic tissue. However, the difference in expression is not significant to conclude the performance of this marker in metastasis of PCa. The same can be said for LRP8, as the trend of its expression profile is analogous to the previous two genes albeit at very low expression levels.

The gene FOXC1 is highly expressed in the metastatic tissue when compared to the primary tumour and normal tissue; this is a contrast to Figure 4.1. Additionally, the survival curve (Figure 4.2) indicates that a high expression of FOXC1 is a good prognosis for metastasis. However, FOXC1 is known to be involved in progression of prostate cancer (Han *et al.*, 2017), several studies have shown it to be highly expressed in prostate cancer tissue and cell lines and even more so in androgen independent prostate cancer and metastasis (van Der Heul-Nieuwenhuijsen *et al.*, 2009). Thus, hormone deprivation therapy for prostate cancer based on androgen may not be a good therapy for individuals with high expression of this gene.

4.6 Conclusion

The study investigated the potential of hsa-miR-5698 target genes as molecular signature markers of prostate cancer prognosis. The prognostic value of the target genes was investigated in terms of log-rank survival, recurrence and metastasis via the online survival analysis tool SurvExpress. The prognostic value of the target genes in terms of log-rank relapse free survival and death analysis of the via ProGgene tool. Results indicated that from the set of six markers studied, a high expression of FOXC1 and LRP8 indicated a poor prognosis for PCa in terms of relapse free survival. A high expression of CDKN1A and ELK1 indicated a poor prognosis for relapse free survival. There was no significant difference between high and low expression for the genes BIRC2 and CTNND1 for relapse free survival (Figure 7.1-7.3 Appendix A).

In terms of overall survival, it was found that high expression FOXC1 leads to poor prognosis for overall survival of the disease. High expression of CTNND1, ELK1 BIRC2 and LRP8 showed a good prognosis for patient outcome in prostate cancer for overall survival (Figure 7.4 – 7.6 Appendix A). In terms of metastasis, it was found that low expression of CDKN1A, CTNND1, BIRC2 and FOXC1 indicates poor prognosis of prostate cancer.

From the data obtained, it can be observed that these markers behave differently for the three outcomes examined. The Human Cancer Metastasis database was used to further investigate the expression profiles of the genes in bone, lymph node and distant site metastasis in prostate cancer. It was found that the expression profiles of CDKN1A and CTNND1 are significantly lower than in normal prostate tissue and even primary tumour, indicating that indeed a low expression of these genes results in a poor outcome in prostate cancer metastasis. However, for the genes ELK1, BIRC2 and LRP8, we see no significant

difference in their expression profiles among the three conditions examined. Whilst FOXC1 is highly expressed in primary tumour and even more so in the metastasized sites. Thus, it may be concluded that in PCa, high expression of FOXC1 is a poor marker for metastasis.

We see from the results obtained that it is difficult to make strong conclusions about the prognostic roles of the targets of hsa-miR-5698 because of their variance in performance. Thus, a number of considerations may come in handy when employing these biomarkers in the management of PCa. One such way would be the combined use of clinical factors with a panel of several of the molecular markers during the early stages of the disease such as CDKN1A and FOXC1. There should be continued observation of these markers after any treatments as indications of relapse and subsequently metastasis.

The variance in performance of these biomarkers across the cohorts used in the studies also demonstrates the need for a personalized approach to prostate cancer management and care. Especially at the screening and diagnostic level as the one-size-fits all PSA screening approach has led to over screening in low risk individuals (Liss *et al.*, 2015).

The differential expression and presence of the protein products of these genes in body fluids could provide an easy means of monitoring PCa, therefore we proceeded to monitor the expression of these proteins in cancer and non-cancer cell lines using a molecular approach in the following chapter.

4.7 References

1. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaranga R, Alberto Rodriguez-Barrientos, José G. Tamez-Peña, Victor Treviño (2013) SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS ONE* 8(9): e74250. doi:10.1371/journal.pone.0074250.
2. Albertsen, P.C., Hanley, J.A., Barrows, G.H., Penson, D.F., Kowalczyk, P.D., Sanders, M.M, Fine, J. (2005). Prostate cancer and the Will Rogers phenomenon. *Journal of the National Cancer Institute*, 97 (17):1248-53.
3. American Society of Clinical Oncology. (2018). Prostate Cancer Statistics. <https://www.cancer.net/cancer-types/prostate-cancer/statistics>.
4. Baselga, J and Arteaga, C.L. (2005). Critical update and emerging trends in epidermal growth factor receptor targeting in cancer. *Journal of Clinical Oncology*, 23:2445–59.
5. Batlle, E., Sancho, E., Franci, C.. (2003). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol* 2000; 2: 84–89.
6. Bello-DeOcampo, D and Tindall, D.J. (2003). TGF- β / Smad Signaling in Prostate Cancer. *Current Drug Targets*, 4(3).
7. Bianco, F.J., Jr, Scardino, P.T., Eastham, J.A. (2005). Radical prostatectomy: long-term cancer control and recovery of sexual and urinary function (“trifecta”) *Urology*, 66:83–94.
8. Buhmeida, A., Pyrhönen, S., Laato, M., Collan, Y. (2006). Prognostic factors in prostate cancer. *Diagnostic Pathology*, 1(4). <http://doi.org/10.1186/1746-1596-1-4>.

9. Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P., Karthaus, H.F., Schalken, J.A., Debruyne, F.M., Ru, N., Isaacs, W.B. (1999). DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Research* 1; 59(23):5975-9.

10. Cao, Z., Kyprianou, N. (2015). Mechanisms navigating the TGF- β pathway in prostate cancer. *Asian Journal of Urology*, 2(1) 11-18.
11. Caram, E.M.V., Skolarus, T.A., Cooney, K.A., (2016). Limitations of Prostate-specific Antigen Testing after a Prostate Cancer Diagnosis. *European Urology*, 70 (2) 209 – 210.
12. D'Amico, A.V., Moul, J.W., Carroll, P.R., Sun, L., Lubeck, D., Chen, M.H. (2003). Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *Journal of the National Cancer Institute*, 95:1376–83.
13. D'Amico, A.V., Whittington, R., Malkowicz, S.B., Schultz, D., Blank, K., Broderick, G.A., Tomaszewski, J.E., Renshaw, A.A., Kaplan, I., Beard, C.J., Wein, A. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*, 280 (11):969-74.
14. de Kruijf, E. M., Dekker, T.J., Hawinkels, I,J,A,C., Putter, H., Smit, V.T.H.B., Kroep, J.R., Kuppen, P.J.K., van de Velde, C.J.H., ten Dijke, P., Tollenaar, R.A.E.M., Mesker, W.E. (2013). The prognostic role of TGF- β signalling pathway in breast cancer patients, *Annals of Oncology*, 24 (2) 384–390, <https://doi.org/10.1093/annonc/mds333>.
15. Ding, Z., Wu, C.J., Chu, G.C, *et al.* (2011). SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, 470:269–73.
16. Dos Reis, S. T., Pontes-Júnior, J., Antunes, A. A., de Sousa-Canavez, J. M., Abe, D. K., da Cruz, J. A. S., Leite, K. R. M. (2011). Tgf- β 1 expression as a biomarker of

- poor prognosis in prostate cancer. *Clinics*, 66 (7), 1143–1147.
<http://doi.org/10.1590/S1807-59322011000700004>.
17. Elian, F. A., Yan, E., Walter, M. A. (2018). FOXC1, the new player in the cancer sandbox. *Oncotarget*, 9 (8), 8165–8178.
<http://doi.org/10.18632/oncotarget.22742>.
18. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. FDA-NIH Biomarker Working Group. 2016 Jan 28 [Updated 2017 Sep 25]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK338449/> Co-published by National Institutes of Health (US), Bethesda (MD).
19. Fizazi, K., Michael Carducci., Matthew Smith., Ronaldo Damião Janet Brown., Lawrence Karsh., Piotr Milecki., Neal Shore., Michael Rader., Huei Wang., Qi Jiang., Sylvia Tadros., Roger Dansey., Carsten Goess. (2011). Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study. *The Lancet*, 377 (9768) 813-822.
20. Fokas E., Engenhardt-Cabillic R., Daniilidis K., Rose F., An H.X. (2007). Metastasis: the seed and soil theory gains identity. *Cancer Metastasis Review*, 26:705–715.
21. Glass, T.R., Tangen, C.M., Crawford, E.D., Thompson, I. (2003). Metastatic Carcinoma of the Prostate: identifying prognostic Groups using recursive partitioning. *The Journal of Urology*, 169, 164–169.
22. Gleason DF, Mellinger GT. (1974). Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol*. 1974 Jan; 111(1):58-64.

23. Goswami, C.P and Nakshatri, H. (2013). PROGgene: gene expression based survival analysis web application for multiple cancers. *Journal of Clinical Bioinformatics* 20133:22. <https://doi.org/10.1186/2043-9113-3-22>.
24. Gravis, G., Boher, J.M., Joly, F., Oudard, S., Albiges, L., Priou, F., Latorzeff, I., Delva, R., Krakowski, I., Laguerre, B., Rolland, F., Theodore, C., Deplanque, G., Ferrero, J.M., Pouessel, D., Mourey, L., Beuzebec, P., Habibian, M., Soulie, M., Fizazi, K. (2015). Androgen deprivation therapy (ADT) plus docetaxel (D) versus ADT alone for hormone-naïve metastatic prostate cancer (PCa): Long-term analysis of the GETUG-AFU 15 phase III trial. *Journal of Clinical Oncology*, 33(7).
25. Zheng, G., Ma, Y., Zou, Y., Yin, A., Li, W., Dong, D. (2017). HCMDB: The Human Cancer Metastasis Database. *Nucleic Acids Research*, 46 (D1) D950–D955, <https://doi.org/10.1093/nar/gkx1008>.
26. Han B, Qu Y, Jin Y, Yu Y, Deng N, Wawrowsky K, Zhang X, Li N, Bose S, Wang Q, Sakkiah S, Abrol R, Jensen TW, Berman BP, Tanaka H, Johnson J, Gao B, Hao J, Liu Z, Buttyan R, Ray PS, Hung MC, Giuliano AE, Cui X. (2015). FOXC1 Activates Smoothed-Independent Hedgehog Signaling in Basal-like Breast Cancer. *Cell Rep*, 13(5):1046-58.
27. Han, B., Bhowmick, N., Qu, Y., Chung, S., Giuliano, A. E., Cui, X. (2017). FOXC1: an emerging marker and therapeutic target for cancer. *Oncogene*, 36(28), 3957–3963. <http://doi.org/10.1038/onc.2017.48>.
28. Hipskind, R.A., Rao, V.N., Mueller, C.G., Reddy, E.S., Nordheim, A. (1991). Ets-related protein Elk-1 is homologous to the c-fos regulatory factor p62TCF. *Nature*; 354:531–534.
29. Huang, W., Chen, Z., Zhang, L. (2015). Interleukin-8 induces expression of FOXC1 to promote transactivation of CXCR1 and CCL2 in hepatocellular carcinoma cell lines and formation of metastases in mice. *Gastroenterology*, 149: 1053–1067.

30. Jin, Y., Han, B., Chen, J., Wiedemeyer, R., Orsulic, S., Bose, S., Zhang, X., Karlan, B.Y., Giuliano, A.E., Cui, Y., Cui, X. (2015). FOXC1 is a critical mediator of EGFR function in human basal-like breast cancer. *Annual Surgical Oncology*, 19:161–169.
31. Jung, B., Staudacher, J. J., Beauchamp, D. (2017). Transforming Growth Factor β Superfamily Signaling in Development of Colorectal Cancer. *Gastroenterology*, 152(1), 36–52. <http://doi.org/10.1053/j.gastro.2016.10.015>.
32. Kalluri, R., Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, 119: 1420–1428.
33. Kan, Z., Jaiswal, B.S., Stinson, J., Janakiraman, V., Bhatt D., Stern, H.M., Yue, P., Haverty, P.M., Bourgon, R., Zheng, J., Moorhead, M., Chaudhuri, S., Tomsho, L.P., Peters, B.A., Pujara, K., Cordes, S., Davis, D.P., Carlton, V.E., Yuan, W., Li, L., Wang, W., Eigenbrot, C., Kaminker, J.S., Eberhard, D.A., Waring, P., Schuster, S.C., Modrusan, Z., Zhang, Z., Stokoe, D., de Sauvage, F.J., Faham, M., Seshagiri, S. (2010). Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*, 466(7308):869-73.
34. Kattan, M.W., Eastham, J.A., Stapleton, A.M., Wheeler, T.M., Scardino, P.T. (1998). A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *Journal of the National Cancer Institute* 90:766–71.
35. Liss, M.A., Chen, H., Hemal, S., Krane, S., Kane, C.J., Xu, J., Kader, A.K. (2015). Impact of family history on prostate cancer mortality in white men undergoing prostate specific antigen based screening. *Journal of Urology*, 193(1):75-9.
36. Lonergan, P. E., and Tindall, D. J. (2011). Androgen receptor signalling in prostate cancer development and progression. *Journal of Carcinogenesis*, 10 (20).
37. Ludwig, J.A., Weinstein, J.N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer*, 5(11):845-56.

38. Maki C.G. (2010) p53 Localization. In: p53. *Molecular Biology Intelligence Unit*, 1. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-8231-5_8
39. Marks, L. S., and Bostwick, D. G. (2008). Prostate Cancer Specificity of PCA3 Gene Testing: Examples from Clinical Practice. *Reviews in Urology*, 10(3), 175–181.
40. Martin, N. E., Mucci, L. A., Loda, M., and DePinho, R. A. (2011). Prognostic Determinants in Prostate Cancer. *Cancer Journal* (Sudbury, Mass.), 17(6), 429–437.
41. Merola, R., Tomao, L., Antenucci, A., Sperduti, I., Sentinelli, S., Masi, S., ... Conti, L. (2015). PCA3 in prostate cancer and tumor aggressiveness detection on 407 high-risk patients: a National Cancer Institute experience. *Journal of Experimental & Clinical Cancer Research* : 34(1), 15.
42. Nalejska, E., Mączyńska, E., & Lewandowska, M. A. (2014). Prognostic and Predictive Biomarkers: Tools in Personalized Oncology. *Molecular Diagnosis & Therapy*, 18(3), 273–284.
43. O'Brate, A., Giannakakou, P. (2003). The importance of p53 location: nuclear or cytoplasmic zip code? *Drug Resist Update* 6(6):313-22.
44. Patki, M., Chari, V., Sivakumaran, S., Gonit, M., Trumbly, R., Ratnam, M. (2013). The ETS domain transcription factor ELK1 directs a critical component of growth signalling by the androgen receptor in prostate cancer cells. *Journal of Biological Chemistry*; 288:11047–11065.
45. Pound, C.R., Partin, A.W., Eisenberger, M.A., Chan, D.W., Pearson, J.D., Walsh, P.C. (1999). Natural history of progression after PSA elevation following radical prostatectomy. *JAMA*, 281:1591–7.
46. Roberts, S.G., Blute, M.L., Bergstralh, E.J., Slezak, J.M., Zincke, H. (2001). PSA doubling time as a predictor of clinical progression after biochemical failure

- following radical prostatectomy for prostate cancer. *Mayo Clinic Procedures*, 76 (6):576–81.
47. Ross, P.L., Scardino, P.T., Kattan, M.W. (2001). A catalogue of prostate cancer nomograms. *Journal of Urology*. 165(5):1562-8.
 48. Schlomm, T., Kirstein, P., Iwers. (2007). Clinical significance of epidermal growth factor receptor protein overexpression and gene copy number gains in prostate cancer. *Clinical Cancer Research* 13:6579–84.
 49. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13 (11), 2498-504.
 50. Shao, N., Chai, Y., Cui, J.Q., Wang, N., Aysola, K., Reddy, E.S., Rao, V.N. (1998). Induction of apoptosis by Elk-1 and DElk-1 proteins. *Oncogene* ;17:527–532.
 51. Shariat, S. F., Kattan, M. W., Vickers, A. J., Karakiewicz, P. I., & Scardino, P. T. (2009). Critical Review of Prostate Cancer Predictive Tools. *Future Oncology*, 5(10), 1555–1584. <http://doi.org/10.2217/fon.09.121>.
 52. Shukla-Dave A., Hricak H., Ishill N., et al. (2009) Prediction of prostate cancer recurrence using magnetic resonance imaging and molecular profiles. *Clinical Cancer Research*. 15:3842–9.
 53. Slatkoff, S., Gamboa, S., Zolotor, A.J., Mounsey, A.L, and Jones, K. (2011). PSA testing: When it’s useful, when it’s not. *Journal of Family Practice*. 60(6): 357–360.
 54. Stark, J.R., Perner, S., Stampfer, M.J., Sinnott, J.A., Finn, S., Eisenstein, A.S., Ma, J., Fiorentino, M., Kurth, T., Loda, M., Giovannucci, E.L., Rubin, M.A., Mucci, L.A. (2009). Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3? *Journal of Clinical Oncology*, 27(21):3459-64.
 55. Tahara, M., Schlumberger, M., Elisei, R., Habra, M.A., Kiyota, N., Paschke, R., Corina E.Dutcus., Taro Hihara., Shannon McGrath., Mark Matijevic., Tadashi

- Kadowaki., Yasuhiro Funahashi., Shermant, S.I. (2017). Exploratory analysis of biomarkers associated with clinical outcomes from the study of lenvatinib in differentiated cancer of the thyroid. *European Journal of Cancer*, 75: 213-221.
56. Tan, M. E., Li, J., Xu, H. E., Melcher, K., & Yong, E. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, 36(1), 3–23. <http://doi.org/10.1038/aps.2014.18>.
57. Tang, Y., Qiao, G., Xu, E., Xuan, Y., Liao, M., & Yin, G. (2017). Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer. *OncoTargets and Therapy*, 10, 4527–4534. <http://doi.org/10.2147/OTT.S142149>.
58. Terada, N., Akamatsu, S., Kobayashi, T., Inoue, T., Ogawa, O., & Antonarakis, E. S. (2017). Prognostic and predictive biomarkers in prostate cancer: latest evidence and clinical implications. *Therapeutic Advances in Medical Oncology*, 9 (8), 565–573.
59. Thompson, I.M., Pauer, D.K., Goodman, P.J., Tangen, C.M., Lucia, M.S., Parnes, H.L., Minasian, L.M., Ford, L.G., Lippman, S.M., Crawford, E.D., Crowley, J.J., Coltman, C.A. Jr. (2004). Prevalence of prostate cancer among men with a prostate-specific antigen level ≥ 4.0 ng per millilitre. *New England Journal of Medicine*, 350 (22):2239-46.
60. Tu, W. H., Thomas, T.Z., Masumori, N., Bhowmick, N.A., Gorska, A.E., Shyr, Y., Kasper, S., Case, T., Roberts, R.L., Shappell, S.B., Moses, H.L., Matusik, R.J. (2003). The loss of TGF-beta signalling promotes prostate cancer metastasis. *Neoplasia*, 5 (3) 267-77.
61. van Der Heul-Nieuwenhuijsen, L., Dits, N.F., Jenster, G. (2009). Gene expression of forkhead transcription factors in the normal and diseased human prostate. *BJU International*, 103; 11 1574-1580, <https://doi.org/10.1111/j.1464-410X.2009.08351.x>.

62. 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.
63. Wei, W., Leng, J., Shao, H., & Wang, W. (2015). High PCA3 scores in urine correlate with poor-prognosis factors in prostate cancer patients. *International Journal of Clinical and Experimental Medicine*, 8(9), 16606–16612.
64. Xu ZY, Ding SM, Zhou L, et al. FOXC1 contributes to microvascular invasion in primary hepatocellular carcinoma via regulating epithelial-mesenchymal transition. *Int J Biol Sci* 2012; 8: 1130–1141.
65. Yang, Z., Jiang, S., Cheng, Y., Li, T., Hu, W., Ma, Z., Yang, Y. (2017). FOXC1 in cancer development and therapy: deciphering its emerging and divergent roles. *Therapeutic Advances in Medical Oncology*, 9(12), 797–816. <http://doi.org/10.1177/1758834017742576>.
66. Zarzynska, J.M. (2014). “Two Faces of TGF-Beta1 in Breast Cancer,” *Mediators of Inflammation*. <https://doi.org/10.1155/2014/141747>.
67. Zhang, X., Zhang B, Gao J, Wang X, Liu Z.(2013). Regulation of the microRNA 200b (miRNA-200b) by transcriptional regulators PEA3 and ELK-1 protein affects expression of Pin1 protein to control anoikis. *Journal of Biological Chemistry* ; 288:32742–32752.

Chapter 5

Molecular Validation of microRNA-target genes as Putative Biomarkers for the Detection and progression of Prostate Cancer.

5.1 Introduction

A highly conserved microRNA will target the mRNA of hundreds of distinct genes to act as a primary regulator of gene expression in many basic cellular processes, including cell proliferation, cell differentiation and apoptosis (Section 1.5). In animals, this is primarily through TNRC6-mediated repression. Thus, there has been increasing evidence that microRNAs can directly regulate specific intracellular mRNA concentrations by RNA degradation, as well as regulate mRNA levels by targeting the cascade of proteins that modulate their downstream transcriptional or post-transcriptional regulation (Park and Shin, 2014; Xu *et al.*, 2016).

In this study, we focus on the putative targets of hsa-miR-5698, which is known to be up-regulated in the prostate adenocarcinoma cell line (LNCaP) and a prostate metastasis cell line (PC3). A functional role has been described for hsa-miR-5698 in transcriptional mis-regulation in cancer (Lombe, 2015). To further investigate the role of hsa-miR-5698, its influence on the mRNA levels of its putative genes CDKN1A, BIRC2, FOXC1, ELK1, CTNND1 and LRP8 as well as their co-regulatory genes (CCND1, TP53, PCNA, CDK2 and CDK4) generated by STRING and IntAct (Sections 3.15 and 31.6) in the prostate cancer cell lines LNCaP and PC3 was analysed using qPCR. This is in a bid to investigate their potential as diagnostic and therapeutic targets in prostate cancer as well to build regulatory networks that the genes may be involved in.

5.1.1 Quantitative real-time PCR (qPCR)

The quantitative real-time polymerase chain reaction (qPCR) analysis of gene expression has rapidly become the most widely used measure of gene expression in many fields of biological research (Kuang *et al.*, 2018). 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 (Higuchi *et al.*, 1993; Williams, 2008).

This method has been used as a common approach to measure the expression of target genes in a wide range of samples from many sources, such as tissues blood, and cultured cells and can be taken advantage of in the study of various pathological states such as cancer. It offers a broad range of advantages over other standard RNA quantification methods such as the Northern blot, and *in situ* hybridization due to its specificity, sensitivity, simplicity, low cost and high-throughput nature (Adamski *et al.*, 2014; Kralik and Ricchi, 2017; Kuang *et al.*, 2018).

5.1.2 Quantification Strategies in qPCR

There are two strategies used to quantify gene expression in qPCR; 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 (Rodriguez-Lazaro and Hernandez, 2014; Tille, 2017). It corresponds to the cycle at which a statistically significant increase in fluorescence is first detected (Tille, 2017). 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; Gotfred-Rasmussen *et al.*, 2016). This reference gene, which is also known as endogenous control, provides a basis for normalizing sample-to-sample differences (Heid *et al.*, 1996). Moreover, qPCR can also provide semi-quantitative results without standards but with controls used as a reference. In this case, the observed results can be expressed as higher or lower multiples with reference to the control. This application of qPCR has been extensively used for gene expressions studies (Bustin *et al.*, 2009).

5.1.3 Aims and objectives

This section of the study aimed to determine the expression of hsa-miR-5698 target genes and their co-expressed genes in the prostate cancer cell lines PC3 and LNCaP using Real-time Polymerase Chain Reactions (qPCR).

Specific Objectives:

1. Growth of a prostate cancer cell line LNCaP (adenocarcinoma) and PC3 (bone metastasis) and mRNA extraction followed by cDNA synthesis.
2. Molecular expression profiling of the six hsa-miR-5698 target genes and their co-expressed genes via qPCR in LNCaP and PC3 cell lines.
3. Identify which target genes in the examined set are related to two stages of PCa.
4. Use the information to understand the regulatory networks that hsa-miR-5698 and its targets are involved in prostate cancer (Chapter 3).

5.2 Materials and methods

All the reagents used in the study as well as their suppliers are listed alphabetically in Table 5.1.

Table 5.1: General reagents and suppliers

Reagent	Supplier
Kaighn's Modification of Ham's F-12 Medium (F-12K)	Lonza
Roswell Park Memorial Institute Medium (RPMI) 1640	Lonza
Dimethyl Sulphoxide (DMSO)	Sigma
Fetal Bovine Serum (FBS)	Lonza
Phosphate Saline Buffer (PBS)	Lonza
Trypsin EDTA	Lonza
β -mercaptoethanol	Sigma
KAPA SYBR [®] FAST qRT-PCR Kit	KAPA Biosystems
RNeasy Mini Kit (RNA extraction)	Qiagen
Nuclease free dH ₂ O	Merck
Transcriptor First Strand cDNA synthesis kit	Roche diagnostics
PenStrep (antibiotic)	Lonza

5.2.1. Cell culture

The cell lines used in the study as well as the media used to maintain them are depicted in Table 5.2. The LNCaP (passage 5) cell line was purchased from American Type Culture Collection (ATCC). The PNT1a (passage 7) and PC3 (passage 4) cell lines were obtained from Luiz Zerbini of the International Centre for Genetic Engineering and Biotechnology (ICGEB) Cape Town. All three cell lines are epithelial and adherent.

Table 5.2: Cell lines used to investigate gene expression in prostate cancer.

Cell line	Tissue	Description	Complete growth medium
LNCaP	Prostate	AR-sensitive prostate adenocarcinoma	RMPI, PenStrep, FBS 10 %
PC3	Prostate	Bone metastasis	F-12K, PenStrep, FBS 10 %
PNT1a (Normalizer)	Prostate	Normal prostate epithelium	RPMI, PenStrep, FBS 10%

5.2.2. Start-up of cell culture from frozen cells

The frozen cryovials were held under 25 °C 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 media was added (Table 5.2). 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-suspended in fresh culture media and the suspension transferred to a 25 cm² flask (T25). The flask was then incubated in a humidified incubator at 37 °C with 5 % CO₂ for 24 hours after which the media was checked for contamination. After an additional 24 hours of culturing, the flask was viewed under a Nikon TMS microscope at 200X magnification to check if the cells had adhered to the flask. The media was removed and replaced with fresh culture media to remove any traces of DMSO left over from the cryopreservation media.

5.2.3. Maintaining the cell lines

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

5.2.4. Sub-cultivation and trypsinization for mRNA extraction

To sub-cultivate (passage) the cultures, old media was aspirated with a sterile Pasteur pipette and discarded. The culture was then washed with 3 ml 1X Phosphate Buffered Saline (PBS) (Table 5.1) pre-warmed at 37 °C which was then discarded. One millilitre of 1.25 % trypsin EDTA was added to the culture and the flask incubated at 37 °C 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 media was added to deactivate the trypsin. The suspended cells were collected to a pellet by centrifugation at 2039 x g to be used for RNA extraction.

5.2.5. Extraction of RNA

The procedure for the extraction of RNA was performed on the LNCaP and PC3 cell lines according to the manufacturer's instructions (RNeasy Mini Kit Qiagen). For the RNA extraction, the confluent cell lines were harvested as per Section 5.2.4, at a concentration of 4.8×10^6 cells/ml. The cells were transferred to an Eppendorf Tube® and centrifuged at $2000 \times g$ for 5 minutes at 4 °C using an Eppendorf 5417R bench top centrifuge. Thereafter, the supernatant was removed and the cells in the pellet lysed by adding 350 µL of RLT lysis buffer containing 3.5 µL of a β-mercaptoethanol solution. The resulting lysate was homogenised by passing through a 20 gauge needle (0.9 mm fitted to an RNase-free syringe) five times. A volume of 250 µL of 70 % ethanol was added to the lysate and 700 µL of the resulting solution added to a RNeasy Mini spin column placed in a 2 ml collection tube. This was centrifuged for 15 s at $8000 \times g$ and the flow-through discarded. A volume of 500 µL of RPE buffer was added to the

RNeasy spin column which was centrifuged for 15 s at 8000 x g to wash the membrane, the flow-through was discarded. This step was repeated. The RNeasy spin column was placed in a new 2 ml collection tube and centrifuged at 12 000 x g for 1 min to dry the membrane. The RNeasy spin column was placed in a new 1.5 ml collection tube and a volume of 40 µL RNase-free water added directly to the spin column membrane, the tube centrifuged for 1 min at 8000 x g to elute the RNA. The concentration and quality of RNA was assessed using the Nanodrop ND-1000 spectrometer (ThermoScientific) and all the RNA samples were stored at -80 °C.

5.2.6. Primer design

Gene specific primers were designed for qPCR and their sequences are shown in Table 5.3. Each primer was designed to be 20 bp long using the NCBI Primer-BLAST algorithm accessible at <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>.

The oligonucleotide sequences were sent for synthesis to Inqaba biotech <http://www.inqababiotec.co.za/>. The primers were delivered as a lyophilized pellet and a 10 µM working stock solution was prepared by re-suspending the pellet in 1X TE buffer (10 mM Tris, pH 7.5 to 8.0, 1 mM EDTA), the primers were stored at -20 °C.

Table 5.3: Primer sequences for PCR amplification of cDNA

Primer	FWD 5' - 3'	REV 5' - 3'
CDKN1A	TGCCCGCCTTTCTTTTTGAG	AGTTTGAAGCTGCAGTGAGC
CCND1	AAAAACCGTCCACAGCAGAG	AAGCTCCAAAAAGGCAGCAC
TP53	AGCTCGCGGTTGTTTCATTC	ACTGGCGCTGTGTGTAATG
LRP8	CAACATGCAACTACCCATGC	GATTTCCCTCACCCCAAAT
PCNA	AACCGCGTTCGAAATACAGC	TCATTGCCGGCGCATTITAG
BIRC2	TTGGGCTTGTTGTGTTGGTG	CGCTGTCTTTCTGCAAACAC
CDK2	TGTTGGCACACTGATTCAGC	TAAATGGGCAGCAGGTGTTT
CDK4	TGCAAGGCATGTGTCATGTG	AGGCCCTGCAATAGAAAACG
CTNND1	GCTGCCAGATCAGTTTGCA	GCCAAGGTGCTGAGAAAGAC
ELK1	CAGACCCCAGCTACTTCTCG	GAGACAGGAGCCACAAGAGG
FOXC1	CTCCCCTCTCTTGCCTTCTT	CGTCAGGTTTTGGGAACACT
GAPDH	ACCCACTCCTCCACCTTTG	CTCTTGCTCTTGCTGGG
HPRT-1	TGCTCGAGATGTGATGAAGG	TCCCCTGTTGACTGGTCATT

5.2.7. Reverse transcription of mRNA to cDNA

The cDNA was synthesized using the Transcriptor First Strand cDNA synthesis kit from Roche diagnostics, 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 5.4 in sterile, nuclease-free, thin walled PCR tubes to a final volume of 13 μ L. The tubes were then incubated at 65 °C for 10 minutes,

after which the cDNA synthesis reagents in Table 5.5 were added to make a final volume of 20 μL .

Table 5.4: Reagents and components for cDNA synthesis

Component/Reagent	Volume	Final Concentration
RNA	Variable	1 μg
Oligo dT primer	1 μL	2.5 μM
PCR grade water	Variable	
Transcriptor Reverse Transcriptase Reaction Buffer	4 μL	1 X (8mM MgCl_2)
Protector RNase Inhibitor	0.5 μL	20 U
Deoxynucleotide Mix	2 μL	1 mM
Transcriptor Reverse Transcriptase	0.5 μL	10 U
Final volume	20 μL	

The reaction was incubated at 55 $^{\circ}\text{C}$ for 30 min followed by an transcriptase inactivation step of 5 minutes at 85 $^{\circ}\text{C}$. The concentration of the synthesised cDNA was determined with a NanoDrop Spectrophotometer ND1000. The cDNA was diluted to a final working concentration of 250 ng.

5.2.8 Analysis of gene expression profiles of the genes in cancer and control cell lines using qPCR

Expression profiles of the genes were analysed via qPCR in the LNCaP and PC3 cell lines as well as the PNT1a cell line, which was used as the normalizer. The house keeping genes GAPDH and HPRT-1 were used as calibrators. All reactions were performed on the LightCycler® 480 System (Roche diagnostics) instrument. The reactions were prepared as outlined in Table 5.5.

Table 5.5: Reagents for a standard qPCR reaction

Reagents	Final Concentration
SYBR Green Master Mix (10 X)	1 X
Forward Primer	1 μ M
Reverse Primer	1 μ M
cDNA	25 ng
PCR Grade dH ₂ O	Variable to make 20 μ L
Final Volume	20 μ L

The qPCR reactions were performed by analysis of each of the genes in each cell line. In addition, reactions for the reference housekeeping genes and a no-template control (water) were also set up for each cell line. A 18 μ L aliquot of

reaction master-mix was pipetted into each well of a 96 well plate and a 2 μ L aliquot (250 ng) of cDNA from each cancer cell line was then added as the qPCR template to each well respectively. The 96 well plates were sealed with clear sealing film and a qPCR run set up on the LightCycler[®] 480 instrument according to the parameters in Table 5.6 and 5.7. The evaluating parameters selected for data analysis were fluorescence ($d[F1]/dT$), melting temperature (T_m) and crossing point (C_p). The Second Derivative Maximum algorithm was employed for C_p determination where C_p 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 250 ng to 0.025 ng. The PCR efficiencies were calculated using the REST[®] software and all threshold cycle (C_t) values were taken into consideration according to the following equation: $E=10[-1/slope]$ (Pfaffl 2002).

Table 5.6: Cycling Protocol for the qPCR

Programme	Cycles	Analysis Mode
Pre-incubation	1	None
Amplification	45	Quantification
Melting Curve	1	Melting Curve
Cooling	1	None

Table 5.7: Specific LightCycler® 480 parameters for the qPCR

Programme Name	Target Temp (°C)	Acquisition Mode	Hold (hh:mm:ss)
Pre-incubation	95	None	00:03:00
Amplification	95	None	00:00:10
	Primer Dependent 50 - 67	One	00:00:20
Melting curve	95	None	00:00:05
	65	None	00:01:00
	97	Continuous	5-10 acquisitions/ °C
Cooling	40	None	00:00:10

5.3 Results and Discussion

The expression profiles of a set of 11 genes comprising six putative translational targets of hsa-miR-5698 and their five co-expressed genes were analysed using qPCR, in the PNT1A, LNCaP and PC3 cell lines. The results of the analysis are depicted in Figure 5.1. This discussion will focus on and highlight those genes whose expression profiles shows significant changes between the cell lines.

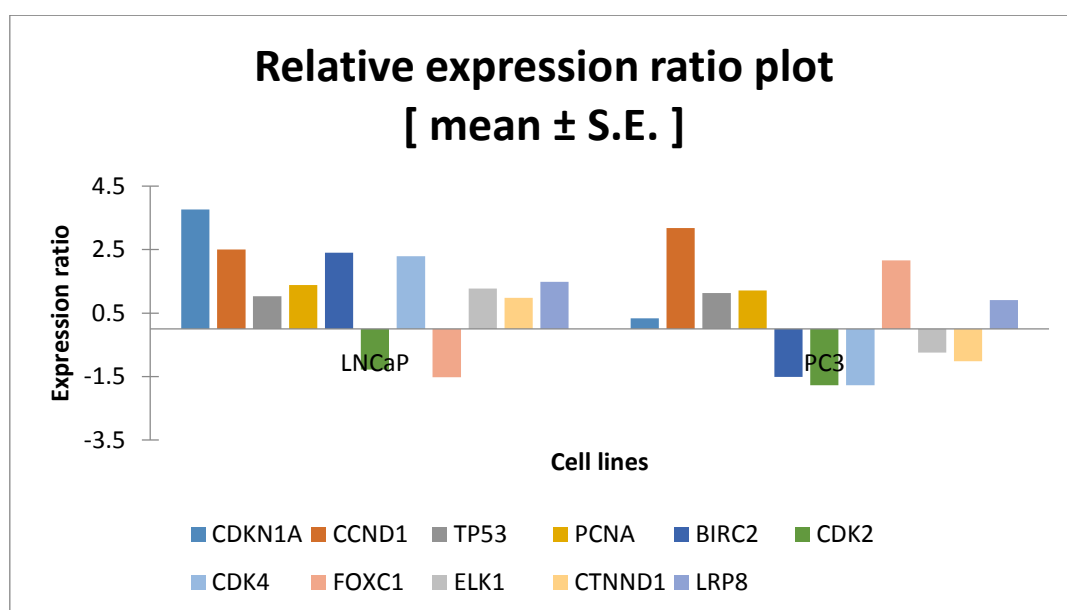


Figure 5.1: Relative expression ratio plot of the targets of hsa-miR-5698 targets and their co-expressed genes in the prostate cancer cell lines LNCaP and PC3.

The qPCR results showed an up-regulation of CDKN1A (p21), ELK1, CTNND1, BIRC2 and CDK4 in LNCaP, and their down-regulation in PC3. There was a down-regulation of FOXC1 in LNCaP and an up-regulation in PC3. From Figure 2.4, it was seen that the expression profile of CDKN1A showed a moderate negative linear relationship to the expression profile of hsa-miR-5698 in PC3 (-0,60) and a highly negative linear relationship in LNCaP (-0,95). The gene p21 was found to be the most robust target of hsa-miR-5698 (Lombe, 2015). This was based on various criteria used in microRNA target prediction algorithms, such a highly conserved seed region and microRNA-recognition elements (MREs). Thus, the gene scored highly as a target in four databases, namely, TargetScanHuman, miRDB, miRBase as well as DIANA Tools.

The protein product of this gene is a potent cyclin-dependent kinase inhibitor, which inhibits the activity of cyclin-CDK2, -CDK1, and -CDK4/6 complexes, and thus functions as a regulator of cell cycle progression at G1 and S phase (Chuxia *et al.*, 1995; Gartel and Radhakrishnan, 2005; Li *et al.*, 2017). Many studies examining the role of p21 on CDK2 activity in cell lines have shown it to have a bifurcate role following mitosis, such that, cells with high p21 levels enter a G0/quiescent state, whilst those with low p21 levels exhibit increased proliferation (Spencer *et al.*, 2013; Li *et al.*, 2017). This may perhaps explain the result obtained in the qPCR, where the metastasized cell line exhibits lower expression profiles of the gene (Figure 5.1).

Place *et al.*, 2012 as well as Li *et al.*, 2017 demonstrated the effect of microRNA action on p21 in prostate cancer both *in vitro* and *in vivo*. The latter study demonstrated that an overexpression of endogenous hsa-miR-3619-5p induced p21 expression and down-regulated cyclin D1-CDK4/6 levels, and consequently inhibited prostate cancer cell proliferation. Both studies cited microRNA

activation of gene expression by the targeting of promoter elements as a possible mechanism of action in their experiments.

In the correlation analysis performed in Section 2.3.3, it was seen that the expression profile of FOXC1 showed a strong positive linear relationship to hsa-miR-5698 in PC3 (0,84) (Figure 2.4), this suggests that as the expression profile of hsa-miR-5698 decreases, the expression profile of FOXC1 increases in PC3. In contrast, there is a weak negative linear relationship in the LNCaP cell line (-0,25). The product of FOXC1 is a transcription factor which has various roles in cancer, including the promotion of cell proliferation and contribution to tumour angiogenesis (Li *et al.*, 2017; Elia *et al.*, 2017).

Additionally, several studies have shown a positive correlation between FOXC1 expression and tumour stage, tumour size, stromal invasion, and lymph node metastasis in gastric cancer, melanoma, cervical carcinoma and breast cancer (Ray *et al.*, 2010; Ray *et al.*, 2011; Xu *et al.*, 2016; Wang *et al.*, 2016; Huang *et al.*, 2017). However, overexpression of FOXC1 has been known to inhibit invasive progression in prostate cancer and there is no significant evidence between FOXC1 expression and clinicopathological features of the disease (van Der Heul-Nieuwenhuijsen *et al.*, 2009; Yang *et al.*, 2017).

The BIRC2 gene encodes the Baculoviral IAP repeat-containing protein 2. It is a member of the inhibitor of apoptosis family that inhibit apoptosis by interfering with the activation of caspases (Liston *et al.*, 1996; Yang *et al.*, 2016). This gene has been found to be deregulated in many cancer cells including prostate cancer and thus, may represent a potential target for therapy (Parajuli *et al.*, 2014). Because of its role as an apoptosis inhibitor, its down regulation has been linked to a possible contribution to prostate cancer aggression (Luk *et al.*, 2014). This is a possible explanation of what is observed in the qPCR result (Figure 5.1), where

there is overexpression of BIRC2 in the metastasized cell line PC3 when compared to the localized adenocarcinoma LNCaP.

Animal models have demonstrated that rno-miR-29 is a direct target of BIRC2. Its overexpression reduces BIRC2 mRNA and protein levels leading to an increase in apoptosis, whilst its down regulation plays an apoptosis inducing role in neuroblastoma (Huang *et al.*, 2015). In this study, there is a weak negative correlation between the expression profiles of hsa-miR-5698 and BIRC2 in both cell lines Table 2.1. However, this was a predicted relationship and would need to be validated via a luciferase assay as well as microRNA transfection mimic and inhibition studies.

The gene ELK1 codes for a transcription factor that is known to be involved in a number of functions including regulation of cell proliferation, cell cycle and apoptosis (Shao *et al.*, 1998; Zhang *et al.*, 2013). The androgen receptor, a key driver of tumour progression, is a co-activator of ELK1 in PCa and silencing it retards the proliferation of AR- positive cells (Patki *et al.*, 2013). This could be an explanation for what can be seen in Figure 5.1 where the LNCaP cell line, which is androgen sensitive, expresses it more when compared to the PC3 cell line which is AR-independent. There has been very little evidence to describe the role of this gene in prostate cancer apart from the referenced publication. This study used qPCR to support the previous publication. However, a larger panel of cell lines as well as investigations in tissue biopsies should be used to further validate these findings.

The gene CTNND1 (p120) encodes a member of the Armadillo (Arm) family of proteins, which function in cell-cell adhesion and signal transduction. It functions to stabilize the E-cadherin-based complex that controls tissue integrity at cell-cell junctions (Dohn and Reynolds, 2011; Schackmann *et al.*, 2013). Loss or

functional inactivation of this gene is a cardinal event in the acquisition of migration and invasion properties to a cell, which is strongly linked to the progression towards metastasis (Vleminckx *et al.*, 1991; Jeanes *et al.*, 2008; Kalluri and Weinberg, 2009). This could explain the result seen in Figure 5.1, where the gene is down-regulated in PC3. Recently, a study by Tang and co-workers examined the role of this gene in hepatocellular carcinoma (HCC) and found that it promoted cell proliferation, migration and invasion *in vitro* and promoted HCC cell tumour formation and metastasis *in vivo* (Tang *et al.*, 2016). However, a role for this gene in prostate cancer has not yet been elucidated and this study is the first to attribute differential expression of the gene between two stages of PCa *in vitro* (Tang *et al.*, 2016).

In another study, Wu *et al.*, 2016, predicted that this gene is a target of hsa-miR-409-3p in osteosarcoma and examined the effect of the microRNA on the gene in the disease. They found that overexpression of miR-409-3p in osteosarcoma cells (U2OS) inhibited cell migration and invasion via the repression of the gene (Wu *et al.*, 2016). Thus, validation of CTNND1 as a target of hsa-miR-5698 as well microRNA transfection investigations are cardinal to the understanding and elucidation of the role of the gene in PCa.

5.4 Conclusion

The expression levels of hsa-miR-5698 targets and their co-expressed genes were determined in the prostate cancer cell lines PC3 and LNCaP using qPCR. It was found that there was an up-regulation of CDKN1A (p21), BIRC2 and CDK4 in

LNCaP, and their down-regulation in PC3. There was also a down-regulation of FOXC1 in LNCaP and an up-regulation in PC3.

Over the past few years, there have been a number of genes found to be involved in prostate cancer progression including two genes examined in this study, BIRC2 and CDKN1A (Hughes *et al.*, 2005; Parajuli *et al.*, 2014). However, there has been no linkage of the genes to various stages of the disease. Additionally, there are very few studies that investigate the dual roles of microRNAs and their targets in prostate cancer as well as its progression. This study indicated significant differences in the expression profiles of hsa-miR-5698 target genes between the two prostate cancer cell lines which correspond to two different stages of the disease. It was found that CDKN1A (p21), ELK1, CTNND1, BIRC2 and FOXC1 are able to distinguish between LNCaP and PC3 cell lines, with the FOXC1 gene being the only one in the set that was overexpressed in PC3.

However, more cell lines and patient samples would need to be evaluated to establish the specificity of the expression of these genes in the stages of PCa. These markers hold great potential in the improvement of screening predictive accuracy or prognostics as well as treatment outcomes in PCa.

5.4 References

1. Adamski, M. G., Gumann, P., Baird, A. E. (2014). A method for quantitative analysis of standard and high-throughput qPCR expression data based on input sample quantity. *PLoS One*, 9(8), e103917. doi:10.1371/journal.pone.0103917.
2. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4) pp. 611-22.
3. Chuxia, D., Zhang, Pumin; Harper, J. Wade; Elledge, Stephen J.; Leder, Philip (1995). "Mice Lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control". *Cell*. Elsevier BV. 82 (4) pp. 675–684. doi:10.1016/0092-8674(95)90039-x.
4. Dohn, M.R., Reynolds, A.B. (2011). CTNND1 (catenin (cadherin-associated protein), delta 1). *Atlas Genetics Cytogenetics Oncology Haematology*, 15 (2) pp. 128-131. doi:10.3389/fmicb.2017.00108.
5. Gartel, A.L., Radhakrishnan, S.K. (2005). "Lost in transcription: p21 repression, mechanisms, and consequences". *Cancer Research*, 65 (10) pp. 3980–5. doi:10.1158/0008-5472.CAN-04-3995. PMID 15899785.
6. Gotfred-Rasmussen, H., Lund, M., Enemark, H.L., Erlandsen, M., Petersen E. (2016). Comparison of sensitivity and specificity of 4 methods for detection of *Giardia duodenalis* in feces: immunofluorescence and PCR are superior to microscopy of concentrated iodine-stained samples. *Diagnostic Microbiology and Infectious Disease*, 84 (3) pp. 187-190.

7. Heid, C. A., Stevens, J., Livak, K.J., Williams, P.M. (1996). Real-time quantitative PCR. *Genome Research*, (6) pp. 986-994.
8. Higuchi, R., Fockler, C., Dollinger, G., Watson, R. (1993). Kinetic PCR Analysis – Real-Time Monitoring of DNA Amplification Reactions. *Bio/technology*, (Nature Publishing Company). 11. 1026-30. 10.1038/nbt0993-1026.
9. Holland, M.J. (2002). Transcript abundance in yeast varies over six orders of magnitude. *Journal of Biological Chemistry*, 277(17), pp. 14363–6. doi: 10.1074/jbc.C200101200.
10. Huang, L., Huang, Z., Fan, Y. (2017). FOXC1 promotes proliferation and epithelial-mesenchymal transition in cervical carcinoma through the PI3K-AKT signal pathway. *American Journal of Translational Research*, 9 pp. 1297–1306.
11. Huang, L.G., Li, J.P., Pang, X.M., Chen, C.Y., Xiang, H.Y., Feng, B.O., Su, S.S., Li, S.H., Zhang, L., Liu, J.L. (2015). MicroRNA-29c Correlates with Neuroprotection Induced by FNS by Targeting Both Birc2 and Bak1 in Neuroblastoma. *CNS Neuroscience & Therapeutics*, 21, (6), pp. 496-503.
12. Hughes, C., Murphy, A., Martin, C., Sheils, O., O'Leary, J. (2005). Molecular pathology of prostate cancer. *Journal of clinical pathology*, 58 (7), pp. 673-84.
13. Jeanes, A., Gottardi, C. J., Yap, A. S. (2008). Cadherins and cancer: how does cadherin dysfunction promote tumour progression? *Oncogene* 27, pp. 6920–6929. doi:10.1038/onc.2008.34
14. Kalluri, R. and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigations*. 119, pp. 1420–1428. doi:10.1172/JCI39104.

15. Kralik, P., Ricchi, M. (2017). A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Frontiers in Microbiology*, 8 (108).
16. Kuang, J., Yan, X., Genders, A.J., Granata, C., Bishop D.J. (2018). An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PLoS ONE* 13 (5), e0196438. <https://doi.org/10.1371/journal.pone.0196438>.
17. Li, S., Wang, C., Yu, X., Wu, H., Hu, J., Wang, S., Ye, Z. (2017). miR-3619-5p inhibits prostate cancer cell growth by activating CDKN1A expression. *Oncology Reports*, 37(1) pp. 241-248. doi: 10.3892/or.2016.5250.
18. Liston, P., Roy, N., Tamai, K., Lefebvre, C., Baird, S., Cherton-Horvat, G., Farahani, R., McLean, M., Ikeda, J.E., MacKenzie, A., Korneluk, R.G. (1996). "Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes". *Nature*. 379 (6563) pp. 349–53. doi:10.1038/379349a0. PMID 8552191.
19. Lombe, C. (2015). Identification of microRNAs as a class of biomarkers for the early diagnosis of prostate cancer: an *in silico* and molecular approach, MSc thesis, University of the Western Cape, Cape Town.
20. Luk, S. U., Xue, H., Cheng, H., Lin, D., Gout, P. W., Fazli, L., Collins, C. C., Gleave, M. E., Wang, Y. (2014). The BIRC6 gene as a novel target for therapy of prostate cancer: dual targeting of inhibitors of apoptosis. *Oncotarget*, 5 (16), pp. 6896-908.
21. Luu, H. N., Lin, H. Y., Sørensen, K. D., Ogunwobi, O. O., Kumar, N., Chornokur, G., Phelan, C., Jones, D., Kidd, L., Batra, J., Yamoah, K., Berglund, A., Rounbehler, R. J., Yang, M., Lee, S. H., Kang, N., Kim, S. J., Park, J. Y., ... Di Pietro, G. (2017). miRNAs associated with prostate cancer risk and progression. *BMC Urology*, 17 (1), 18. doi:10.1186/s12894-017-0206-6.

22. Parajuli, K. R., Zhang, Q., Liu, S., Patel, N. K., Lu, H., Zeng, S. X., Wang, G., Zhang, C., ... You, Z. (2014). Methoxyacetic acid suppresses prostate cancer cell growth by inducing growth arrest and apoptosis. *American journal of clinical and experimental urology*, 2(4), pp. 300-12.
23. Park, J. H., Shin, C. (2014). MicroRNA-directed cleavage of targets: mechanism and experimental approaches. *BMB Reports*, 47 (8), pp. 417-23.
24. Patki, M., Chari, V., Sivakumaran, S., Gonit, M., Trumbly, R., Ratnam, M. (2013). The ETS domain transcription factor ELK1 directs a critical component of growth signalling by the androgen receptor in prostate cancer cells. *Journal of Biological Chemistry*, 288 pp. 11047–11065.
25. Place, R.F., Wang, J., Noonan, E.J., Meyers, R., Manoharan, M., Charisse, K., Duncan, R., Huang, V., Wang, X., Li, L.C. (2012). Formulation of small activating RNA into lipidoid nanoparticles inhibits xenograft prostate tumour growth by inducing p21 expression. *Molecular Therapeutics and Nucleic Acids*. 1:e152012.
26. Q Li., J. Wu, P., Wei, Y. Xu., C. Zhuo, Y. Wang. (2015). Overexpression of Forkhead Box C1 promotes tumour metastasis and indicates poor prognosis in colon cancer via regulating epithelial-mesenchymal transition. *American Journal of Cancer Research*, 5 (6), pp. 2022-2034.
27. Ray, P.S., Bagaria, S.P., Wang, J., Shamonki, J.M., Ye, X., Sim, M.S., Steen, S., Qu, Y., Cui, X., Giuliano, A.E. (2011). Basal-like breast cancer defined by FOXC1 expression offers superior prognostic value: a retrospective immunohistochemical study. *Annual Surgical Oncology*, 18 pp. 3839–47.
28. Ray, P.S., Wang, J., Qu, Y., Sim, M.S., Shamonki, J., Bagaria, S.P., Ye, X., Liu, B., Elashoff, D., Hoon, D.S., Walter, M.A., Martens, J.W., Richardson, A.L, (2010). FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. *Cancer Research*, 70 pp. 3870–76.

29. Rodriguez-Lazaro, D., Hernandez, M. (2014). Encyclopaedia of Food Microbiology. Vol. 2. in Tortorello, M.L. (Ed). London UK: Elsevier, pp. 344-350.
30. Schackmann, .R. C. J., Tenhagen, M., van de Ven R. A. H., Derksen, W. B.P. (2013). p120-catenin in cancer – mechanisms, models and opportunities for intervention. *Journal of Cell Science*, 126 pp. 3515-3525, doi: 10.1242/jcs.134411.
31. Shao, N., Chai, Y.L., Cui, J.Q., Wang, N., Aysola, K., Reddy, S.P., Veena, N.R. (1998). Induction of apoptosis by Elk-1 and Δ Elk-1 proteins. *Oncogene*, 17, pp. 527–532.
32. Spencer, S. L., Cappell, S. D., Tsai, F. C., Overton, K. W., Wang, C. L., Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*, 155 (2), pp. 369-83.
33. Tang, B., Tang, F., Wang, Z., Qi, G., Liang, X., Li, B., Yuan, S., Liu, J., Yu, S., He, S. (2016). Overexpression of CTNND1 in hepatocellular carcinoma promotes carcinous characters through activation of Wnt/ β -catenin signalling. *Journal of Experimental and Clinical Cancer Research*, 35:82. <https://doi.org/10.1186/s13046-016-0344-9>.
34. Tille, P. (2017). Bailey and Scott’s Diagnostic Microbiology. 14th ed. St Louis Missouri: Elsevier, pp. 113- 142.
35. van Der Heul-Nieuwenhuijsen, L., Dits, N.F., Jenster, G. (2009). Gene expression of forkhead transcription factors in the normal and diseased human prostate. *BJU International*, 103, (11) pp. 1574-1580, <https://doi.org/10.1111/j.1464-410X.2009.08351.x>.
36. van Peer, G., Mestdagh, P., Vandesompele, J. (2012). Accurate qPCR gene expression analysis on cell culture lysates. *Scientific Reports* (2) pp. 222-28.
37. Vleminckx, K., Vakaet, L. Jr., Mareel, M., Fiers, W. and van Roy, F. (1991). Genetic manipulation of E-cadherin expression by epithelial tumour cells

- reveals an invasion suppressor role. *Cell* 66, pp. 107–119. doi:10.1016/0092-8674(91)90143-M.
38. Wang, J., Li, L., Liu, S., Zhao, Y., Wang, L., & Du, G. (2016). FOXC1 promotes melanoma by activating MST1R/PI3K/AKT. *Oncotarget*, 7 (51), pp. 84375-84387.
39. Williams, P. M. (2008). The Beginnings of Real-Time PCR. *Clinical Chemistry*, 55 (4), pp. 833–834.
40. Wu, S., Du, X., Wu, M., Du, H., Shi, X., Zhang, T. (2016). MicroRNA-409-3p inhibits osteosarcoma cell migration and invasion by targeting catenin- $\delta 1$. *Gene*, Volume 584, 1 (10), pp. 83-89.
41. Xu, K., Lin, J., Zandi, R., Roth, J. A., Ji, L. (2016). MicroRNA-mediated target mRNA cleavage and 3'-uridylation in human cells. *Scientific Reports*, (6), 30242. doi:10.1038/srep30242.
42. Xu, Y. Y., Tian, J., Hao, Q., Yin, L. R. (2016). MicroRNA-495 downregulates FOXC1 expression to suppress cell growth and migration in endometrial cancer. *Tumour biology. The Journal of the International Society for Onco-developmental Biology and Medicine*, 37(1), pp. 239-51.
43. Xu, Z.Y., Ding, S.M., Zhou, L., Xie, H.Y., Chen, K.J., Zhang, W., Xing, C.Y., Guo, H.J., Zheng, S.S. (2016). FOXC1 contributes to microvascular invasion in primary hepatocellular carcinoma via regulating epithelial-mesenchymal transition. *International Journal of Biological Sciences*, 8(8) pp. 1130-41.
44. Yang, Y., Kelly, P., Shaffer, A. L., Schmitz, R., Yoo, H. M., Liu, X., Huang, D. W., Webster, D., Young, R. M., Nakagawa, M., Ceribelli, M., Wright, G. W., Yang, Y., Zhao, H., Yu, X., Xu, W., Chan, W. C., Jaffe, E. S., Gascoyne, R. D., Campo, E., Rosenwald, A., Ott, G., Delabie, J., Rimsza, L., Staudt, L. M. (2016). Targeting Non-proteolytic Protein Ubiquitination for the Treatment of Diffuse Large B Cell Lymphoma. *Cancer Cell*, 29 (4) pp. 494-507.

45. Yang, Z., Jiang, S., Cheng, Y., Li, T., Hu, W., Ma, Z., Chen, F., ... Yang, Y. (2017). FOXC1 in cancer development and therapy: deciphering its emerging and divergent roles. *Therapeutic advances in medical oncology*, 9(12), pp. 797-816.
46. Yuan, J.S., Reed, A., Chen, F., Stewart, C.N. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7(85) pp. 1-10.
47. Zhang, X., Zhang B, Gao J, Wang X, Liu Z.(2013). Regulation of the microRNA 200b (miRNA-200b) by transcriptional regulators PEA3 and ELK-1 protein affects expression of Pin1 protein to control anoikis. *Journal of Biological Chemistry*, 288 pp. 32742–32752.

Chapter 6

General Discussion and Conclusion

6.1 General discussion

According to the World Health Organization, prostate cancer (PCa) is the second most common cause of cancer in men worldwide and a major cause of cancer-related morbidity and mortality (Bray *et al.*, 2017; IARC, 2018). Additionally, studies have shown that approximately 30 % of PCa patients develop clinical recurrence after radical prostatectomy, and the survival period of this phase of the disease can be limited and variable (Moschini *et al.*, 2016). Thus, the challenge lies in identifying those patients most at risk for relapse, as well as metastasis. Currently, factors such as serum PSA level, Gleason score, and tumour stage are used for diagnosis, prognostication, and treatment decision making (Roberts *et al.*, 2015; Sun *et al.*, 2015).

However, none of these alone or in combination are adequate indicators for accurate clinical decision making in PCa treatment and management. This study aimed to assess the role of the targets of hsa-miR-5698 (a microRNA found to be up-regulated in LNCaP and down-regulated in PC3) in the progression of the disease as well as their role if any, in the aggressiveness of the disease. This was done to determine the usefulness of the genes as a set of dynamic network biomarkers for clinical decision making in PCa.

Using different bioinformatics tools and processes, Lombe (2015) identified a novel microRNA (hsa-miR-5698) associated with prostate cancer. This

microRNA was found to be up-regulated in LNCaP and down-regulated in PC3 in the same study. Bioinformatics algorithms from different databases were used to predict a set of genes targeted by this microRNA. In a bid to assign function to it, the gene ontologies of these genes was performed and it was found that many of them were involved in cell proliferation, apoptosis and migration and invasion properties in prostate cancer. The function of these genes led to the proposal of their usage as putative markers for diagnosis, prognosis and therapeutics in PCa.

An understanding of the relationship in terms of expression profiles between the microRNA and the targets was done using a correlation analysis. This was to determine how the expression of the target genes is related to microRNA up-regulation or down-regulation. An *in silico* analysis was done to identify any possible transcriptional effects of the microRNA on the targets as their prediction was translational. This was in a bid to determine the effects if any, of different types microRNA-gene targeting on PCa progression. A possible prognostic implication of these targets in PCa patient survival, recurrence and metastasis was done using Kaplan-Meier analysis. Finally, the expression profiles of the target genes was examined in LNCaP and PC3 cell lines by qPCR.

This study made use of the advent of high throughput analysis technologies which provide a fast means of genomic analysis for differential gene expression in disease conditions. This allows the prediction and publishing of the protein products of these genes as candidate cancer biomarkers (Dragani *et al.*, 2016; Yuan *et al.*, 2016).

A correlation analysis of the expression profiles of hsa-miR-5698 and its translational targets showed that expression of CDKN1A (p21), CTNND1, BIRC2, LRP8 and ELK1 are negatively correlated to the expression of the

microRNA in PC3 (Figure 2.4). This means that with the down-regulation of this microRNA in the PC3 cell line, the genes are up-regulated. The genes BIRC2 and LRP8 on the other hand exhibited a weak correlation coefficient of - 0.24 and - 0.12 respectively, whilst FOXC1 exhibited a positive correlation (Table 2.1). In the LNCaP cell line, the microRNA was found to be up-regulated and all its targets are negatively correlated with CDKN1A having a strong correlation at -0.95 (Figure 2.4). All the other genes had a moderately strong negative correlation except BIRC2 and ELK1 which had a weak correlation.

The correlation analysis was used as an aid to understand the action of microRNAs on their translational targets and it has been used widely in many publications to understand the relationship between differentially expressed genes and the microRNA acting on them (Wang *et al.*, 2009; Laxman *et al.*, 2015; Wang *et al.*, 2018). However, there are some weaknesses associated to this analysis and they have been extensively documented (Ratner, 2009). One of the shortcomings is that the correlation coefficient interval (-1, +1) is restricted by the individual distributions of the two variables being correlated. Another one is an assumption of linearity is made when performing the correlation. It was thus assumed that the relationship between the microRNA and its target genes is linear. However, this cannot be the case for a number of reasons which will be outlined shortly.

MicroRNAs act to regulate gene expression by binding to the 3' UTR of a mRNA and do either of three things, translational repression, degradation and destabilization of mRNA (Section 1.5) all of these mechanisms result in a down-regulation of gene expression of the particular marker under examination. In this case, a microRNA gene relationship may be said to be

linear. However, this is not the only documented action of microRNAs. They have also been known to up-regulate gene expression in response to the state of the cell and/or in the presence of specific factors (Reimsburg *et al.*, 2018).

Thus, the assumption of linearity cannot be a reliable measure. This can be supported by the result in Figure 5.1, where it can be seen that in LNCaP, all but two of the genes are up-regulated when the microRNA is also up-regulated. Thus, a luciferase assay would need to be performed to first determine which of the putative genes are actually targets of hsa-miR-5698, then microRNA mimic and inhibition studies would be done to monitor how over-expression of the microRNA affects the target expression in real time, in both cell lines.

The transcriptional action of microRNAs on gene expression was also taken into consideration to better understand microRNA regulatory networks. Only two genes (CDKN1A and FOXC1) were found to have sequences complementary to hsa-miR-5698 in their promoter regions. The algorithm Trident was used to identify the microRNA binding sites in genomic DNA. The microRNA sequence was obtained from the MiRbase database <http://www.mirbase.org/> version 22, the promoters of the target genes were extracted from Ensembl genome browser version 94 <https://www.ensembl.org/>.

There have been some recent studies that have examined the formation of triplexes between non-coding microRNAs and DNA. One study by Jenjaroenpun *et al.*, used computational methods to predict triplex target DNA sites for several microRNAs, these sites were then presented in a database called the Triplex Target DNA Site Mapping and Integration (TTSMI; <http://ttsmi.bii.a-star.edu.sg>) (Jenjaroenpun *et al.*, 2015). A more

recent study by Paugh and co-workers used the Trident algorithm to identify several specific binding sites in five genomes including *Homo sapiens* (Paugh *et al.*, 2016).

Thus, the regulatory action of microRNAs on gene promoter regions can be extensively studied via computational methods and several conclusions may be made. However, there is a caveat to the current computational methods available, algorithms such as Trident favour microRNAs with a purine-pyrimidine imbalance. MicroRNAs with a purine or pyrimidine content of 75 % or higher account for over 90 % of the binding sites. In addition to purine-pyrimidine content imbalance being an important determinant of triplex formation, lower than average U content, higher than average G or C content also predicts affinity for double stranded DNA binding (Jenjaroenpun *et al.*, 2015; Paugh *et al.*, 2016). Thus, hsa-miR-5698 conformed to both these conditions (Table 2.1).

To further understand the effect of hsa-miR-5698 and its targets in PCa regulatory networks, novel DNA binding motifs that bind to the promoters of the target genes were investigated via the MEME SUITE database <http://meme-suite.org/tools/meme> version 5. Three novel motifs (Figure 3.5) were discovered on a set of promoters of 11 genes, comprising the six targets of hsa-miR-5698 and five genes which are putatively co-regulated with the target genes. The discovered motifs were putatively matched to a number of known transcription factors involved in various gene ontologies including transcriptional control and DNA damage checkpoints (Table 3.3).

The transcription factors together with hsa-miR-5698 and its targets were used to build regulatory networks involved in the progression of PCa. However, only two transcription factors corresponding to motif-1 and motif-2 were used in the construction of the putative regulatory networks.

Motif 3 and its transcription factor were eliminated in the putative network construction for PCa progression; this is because it was the shortest motif (Figure 3.7) in terms of height, indicating a weakness in its proportion to the negative logarithm of the p-value of the site. The putative regulatory networks constructed identified two FFLs involved in the regulation of PCa progression. The first was the microRNA-FFL, in which hsa-miR-5698 acted as both a repressor and/or activator of both the gene and the TF. The second was the TF-FFL, in which the TF acted as a regulator of the genes as well as hsa-miR-5698.

The positive and negative modulating effects of microRNAs in gene regulatory networks may both contribute to gene expression, and the findings in this study demonstrate that. Additionally, there are a number of studies published that have attempted to build microRNA-gene-TF regulatory networks in cancer for various purposes, such as monitoring cancer progression. In one study conducted in colorectal cancer (CRC), Wang and co-workers used *in silico* prediction data to construct regulatory networks for CRC prognosis identifying two cardinal microRNAs, two genes and one transcription factor having good prediction for survival of the disease (Wang *et al.*, 2017). Sadeghi *et al.*, 2016, performed a similar study in prostate cancer and were able to identify that miR-671-5p, miR-665, miR-663, miR-512-3p and miR-371-5p deregulate STAT3 in the progression of metastasis in PCa (Sadeghi *et al.*, 2016). However, the current study is the only known study to date that has attempted to use *in silico* prediction data to construct regulatory networks for PCa progression. To validate and strengthen the interactions in the network, the targets of the microRNA would have to be experimentally validated via the luciferase assays, microRNA overexpression studies and CHIP-seq assays for TF verification.

The effective prognostic ability of the targets of hsa-miR-5698 was determined for patient metastasis, recurrence and overall survival, using the Kaplan-Meier plots in the various datasets curated in the SurvExpress and ProGgene databases. Results from two datasets indicated that CDKN1A, BIRC2 and FOXC1 could serve as good prognosis markers for metastasis in PCa. Two of these genes, BIRC2 and CDKN1A were under-expressed in PC3 (Figure 5.1), but not LNCaP whilst FOXC1 was over-expressed. This was also reflected in the data from the Lapointe dataset (Figure 4.1) and shows that these genes are capable of distinguishing between high, low and risk patient groups. This coupled with an understanding of the roles of these genes in PCa may provide support for their use as a combined set in the monitoring of metastasis progression in PCa.

The prognostic potential for the set of genes for recurrence in PCa was undertaken in the GSE70769 dataset in the ProGgene database. From the survival curves, p21, FOXC1 and CTNND1 showed significant p-value as predictors for recurrence of PCa, whilst BIRC2, LRP8 and ELK1 were poor predictors of the relapse in the disease in the dataset (Figures 7.1-3, Appendix A). This means that patients with a higher expression of the biomarkers are most likely not to have a recurrence of the disease.

These results were corroborated in the qPCR result where the poor predictor genes are under-expressed in the metastasised cell line PC3. The ability of SurvExpress and ProGgene tools to support results from the qPCR is an indication that the tools are effective for *in silico* analysis of prospective biomarkers. This could perhaps give credence to the further analysis of these genes in PCa management. Further, studies would have to

be conducted in tissue samples as well as in larger population samples than that of the given dataset to monitor the behaviour of the markers.

The prognostic potential for the set of genes for overall survival in PCa was undertaken in the GSE16560 dataset in the PRAD dataset in the TCGA database. From the Kaplan Meier plots, it was seen that only CDKN1A and FOXC1 were good markers for overall survival in PCa, the other genes were poor markers (Figures 7.4-6 Appendix A). However, taking into consideration the length of the course of PCa and the relative survival rate for the localized disease is 98 % in the first five years, the biomarkers would need to undergo further rigorous examinations in larger cohorts. Overall, the prognostic analysis of the genes in PCa showed that an intersection set could be observed for the markers with good prognosis for metastasis, recurrence and overall survival in PCa. This set comprised two genes, p21 and FOXC1.

The differential expression signature of the targets of hsa-miR-5698 and their co-expressed genes was undertaken in two prostate cancer cell lines using qPCR. It was found that six genes out of the set of 11 genes could distinguish between the two stages of the cancer. The genes CDKN1A, BIRC2, CTNND1, ELK1 and CDK4 were significantly down-regulated in PC3 when compared to LNCaP, whilst the gene FOXC1 was down-regulated in LNCaP, but significantly up-regulated in PC3. Thus, these genes may also serve as putative biomarkers for the monitoring of the progression of the disease. However, cell lines merely give a glimpse of a single event unlike tumours which are heterogeneous for the disease and thus, qPCR and western blot analysis on the latter samples would be beneficial in this study.

A comparison of the qPCR result with microarray studies on tumours in various stages of PCa in the Human cancer metastasis database showed a corroboration for some of the genes namely, CDKN1A, CTNND1 and FOXC1 (Figure 4.8). The expression of the other genes showed no significant difference between the primary tumour and metastasis.

6.2 Conclusion

The outcome of the analysis, expression profiling and characterization of hsa-miR-5698 target genes as putative dynamic network biomarkers for prostate cancer: a combined *in silico* and molecular approach, was as follows;

- Based on the Pearson correlation analysis, the expression profiles of CDKN1A, CTNND1, BIRC2, LRP8 and ELK1 were negatively correlated to the expression of hsa-miR-5698 in the PC3 and LNCaP cell lines. The expression of FOXC1 on the other hand was positively correlated to the expression of hsa-miR-5698 in the PC3 cell line, but negatively correlated in the LNCaP cell line.
- In addition to being an instigator of translational control of gene expression for the set of genes under investigation, hsa-miR-5698 was also found to bind to the promoters of CDKN1A and FOXC1 using *in silico* analysis in Chapter 2 indicating possible transcriptional control.
- KEGG pathway analysis in Chapter 3 placed one of the genes of interest (CDKN1A) in a cardinal position for cell cycle progression in the prostate cancer pathway.

- Two statistically significant novel sequence motifs were discovered on the promoters of the set of genes comprising hsa-miR-5698 targets and putative genes in their co-expression networks.
- Two putative TFs corresponding to the discovered motifs were identified and used in the building of putative regulatory networks in PCa progression (Chapter 3).
- The regulatory networks identified two feed forward loops involved in the progression of PCa from adenocarcinoma to the metastatic phase, namely the hsa-miR-5698-FFL and the FOXD1/ TBX2-FFL.
- The prognostic analysis of the set of genes targeted by hsa-miR-5698 showed that the two genes CDKN1A and FOXC1 were good putative indicators for metastasis, recurrence and overall survival in PCa. With the high expression of the CDKN1A offering good prognosis for all three conditions examined and the high expression of FOXC1 indicating a poor prognosis for the same (Chapter 4).
- The qPCR analysis in Chapter 5 resulted in six genes being able to distinguish between the PCa stages in two cell lines PC3 and LNCaP. The genes CDKN1A, BIRC2, CTNND1, ELK1 and CDK4 were significantly down-regulated in PC3 when compared to LNCaP, whilst the gene FOXC1 was down-regulated in LNCaP, but significantly up-regulated in PC3.

6.3 Future Work

Further understanding of the expression patterns of the microRNA deregulation and how that affects gene expression may allow development of novel diagnostic, prognostic and therapeutic strategies involving microRNA augmentation or inhibition in the future. The present study revealed that targets of hsa-miR-5698 as well as genes putatively co-expressed with them in PCa regulatory networks play important roles in the progression of prostate cancer. This was because they demonstrated over or under-expression when comparing two prostate cancer cell lines. Some of the genes also demonstrated their ability as prognosis markers for survival of PCa at different stages.

However, validation that indeed the set of six genes are targeted by hsa-miR-5698 would have to be undertaken via luciferase assays. The effect of microRNA over-expression and under-expression studies should also be done via mimics and inhibitors of hsa-miR-5698. Confirmation of the formation of triplexes between hsa-miR-5698 the promoters of genes CDKN1A and FOXC1 must be done via Electrophoretic Mobility Shift Assay (EMSA) and NMR spectroscopy. Further investigation of these genes in actual tumour samples as well as larger patient cohorts will help to define their potential role as possible diagnostic, prognostic and therapeutic biomarkers in the future.

6. 4 References

1. Ajape, A., Ibrahim, K.O., Fakeye, J., Abiola, O.O. (2010) . An overview of cancer of the prostate diagnosis and management in Nigeria: The experience in a Nigerian tertiary hospital. *Annals of African Medicine*, 9 (3), 113. <http://dx.doi.org/10.4103/1596-3519.68353>.
2. Babb, C., Urban, M., Kielkowski, D., Kellett, P. (2014). Prostate Cancer in South Africa: Pathology Based National Cancer Registry Data (1986- 2006) and Mortality Rates (1997-2009). *Prostate Cancer*, pp. 1-9. <http://dx.doi.org/10.1155/2014/419801>
3. Bray F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Zanetti, R., Ferlay, J., editors (2017). Cancer Incidence in Five Continents, Vol. XI (electronic version). Lyon: International Agency for Research on Cancer. Available from: <http://ci5.iarc.fr>, accessed [2/11/2018].
4. Dragani, T. A., Castells, A., Kulasingam, V., Diamandis, E. P., Earl, H., Iams, W. T., ... Schalken, J. A. (2016). Major milestones in translational oncology. *BMC Medicine*, 14(1), 110. <https://doi.org/10.1186/s12916-016-0654-y>.
5. Hoffman, R.M., Gilliland, F.D., Eley, J.W. (2001). Racial and Ethnic Differences in Advanced-Stage Prostate Cancer: the Prostate Cancer Outcomes Study. *Journal of the National Cancer Institute*, 93 (5), pp. 388-395. <http://dx.doi.org/10.1093/jnci/93.5.388>.
6. International Agency for Research on Cancer. EUCAN: Prostate cancer Factsheet. Lyon, France; 2018.

<http://eco.iarc.fr/eucan/CancerOne.aspx?Cancer=29&Gender=1>. Accessed [30/10/2018].

7. Jenjaroenpun P, Chew CS, Yong TP, Choowongkomon K, Thammasorn W, et al. (2015) The TTSMI database: a catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. *Nucleic Acids Res* 43, pp. 110–116. doi: 10.1093/nar/gku970.
8. Moschini, M., Sharma, V., Zattoni F., Quevedo J.F., Davis B.J., Kwon, E., Karnes, R.J. (2016) 'Natural History of Clinical Recurrence Patterns of Lymph Node–Positive Prostate Cancer After Radical Prostatectomy', *European Urology*. Elsevier, 69 (1), pp. 135–142. doi: 10.1016/j.eururo.2015.03.036.
9. Ratner, B. (2009) 'The correlation coefficient: Its values range between +1/–1, or do they?', *Journal of Targeting, Measurement and Analysis for Marketing*. Palgrave Macmillan UK, 17(2), pp. 139–142. doi: 10.1057/jt.2009.5.
10. Remsburg, C., Konrad, K., Sampilo, N. F., & Song, J. L. (2018). Analysis of microRNA functions. *Methods in Cell Biology*. <https://doi.org/10.1016/BS.MCB.2018.10.005>
11. Roberts, M.J., Chow, C.W., Schirra, H.J., Richards, R., Buck, M., Selth, L.A., Doi, S.A., Samaratunga, H., Perry-Keene, J., Payton D. (2015). Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer. *Prostate* 75 (5), pp. 539–549.
12. Shao Y-H, Demissie K, Shih W.(2009). Contemporary Risk Profile of Prostate Cancer in the United States. (JNCI) *Journal of the National Cancer Institute* 2009; 101(18):1280-3. <http://dx.doi.org/10.1093/inci/djp262>.

13. Shenoy, D., Packianathan, S., Chen, A. M., & Vijayakumar, S. (2016). Do African-American men need separate prostate cancer screening guidelines?. *BMC Urology*, 16(1), 19. doi:10.1186/s12894-016-0137-7.
14. Sun X, Yang Z, Zhang Y, He J, Wang F, Su P, Han J, Song Z, and Fei Y (2015). Prognostic implications of tissue and serum levels of microRNA-128 in human prostate cancer. *International Journal of Clinical Experimental Pathology* 8 (7), 8394–8401.
15. Tindall EA, Monare LR., Petersen, D.C., van Zyl, S., Hardi, R.A., Alpheus M., Segone, A.P., Venter, S.M., Bornman, R., Hayes, V. M. (2014). Clinical presentation of prostate cancer in Black South Africans. *The Prostate* 74 (8). <http://dx.doi.org/10.1002/pros.22806>.
16. Wasike, R.W., Magoha, G.A. (2008). Descriptive case series of patients presenting with cancer of the prostate and their management at Kenyatta National Hospital, Nairobi. *East African Medical Journal*, 84 (9).
17. Wang, L., Zhu, J., Deng, FY. et al. (2018). Correlation analyses revealed global microRNA–mRNA expression associations in human peripheral blood mononuclear cells. *Molecular Genetics Genomics* 293: 95. <https://doi.org/10.1007/s00438-017-1367-4>.
18. Wang, H., Luo, J., Liu, C., Niu, H., Wang, J., Liu, Q., ... Zhang, Q. (2017). Investigating MicroRNA and transcription factor co-regulatory networks in colorectal cancer. *BMC Bioinformatics*, 18(1), 388. <https://doi.org/10.1186/s12859-017-1796-4>
19. Yuan, J., Hegde, P. S., Clynes, R., Foukas, P. G., Harari, A., Kleen, T. O., ... Fox, B. A. (2016). Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *Journal for ImmunoTherapy of Cancer*, 4(1), 3. <https://doi.org/10.1186/s40425-016-0107-3>

Appendix A

Chapter 3

Table 7.1: Cis-Regulatory Module (CRM) detections in prostate cancer. The Transcription factor FOXD3 is highlighted in yellow.

ChromID	CRM start pos	CRM end pos	Gene name	TSS pos	Orientation	Min energy	TFs
X	11041910	11041937	NM_013423	11043752	-1	- 1.04075	LHX3::POU3F2::CDC5::MRF-2::C/EBPbeta
X	11041779	11041889	NM_013423	11043752	-1	- 1.15725	LHX3::POU3F2::CDC5::TBP::MRF-2
X	11441359	11441444	NM_013427:NM_006125	11443478	-1	- 1.34933	LHX3::AFP1::FOXJ2::TFIIA::TBP
X	11441263	11441345	NM_013427:NM_006125	11443478	-1	- 1.16739	LHX3::CDC5::AFP1::FOXJ2::TFIIA
X	11440729	11440902	NM_013427:NM_006125	11443478	-1	- 1.52763	LHX3::TBP::CDC5::POU3F2::HNF-1
X	11438478	11438511	NM_013427:NM_006125	11443478	-1	- 1.10624	LHX3::CART-1::C/EBPgamma::POU1F1::CDC5
X	11064848	11065210	NM_013422	11068255	-1	- 3.66859	LHX3::C/EBPgamma::FOXJ2::C/EBP::STAT5A
X	11064192	11064271	NM_013422	11068255	-1	- 1.02577	LHX3::OCT-1
X	49291939	49292016	NM_007003	49296939	1	-	TBP::POU3F2::HNF-1::STAT5A::OCT-1

						1.55932	
X	49292274	49292536	NM_007003	49296939	1	-	POU3F2::MRF-2::CRX::LHX3::FOXD3 1.86017
X	66543001	66543086	NM_000044	66546894	1	-	LHX3::FOXJ2::POU3F2::STAT5A 1.21695
X	66543308	66543638	NM_000044	66546894	1	-	LHX3::C/EBPgamma::AFP1::POU3F2::POU1F1 2.36525
X	66543654	66543664	NM_000044	66546894	1	-	HNF-1::TEF::AFP1::STAT5A::POU3F2 1.03644
X	66543667	66543678	NM_000044	66546894	1	-	TBP::HNF-1::TEF::AFP1::STAT5A 1.11525
X	66543728	66543732	NM_000044	66546894	1	-	HNF-1::TBP::TEF::NKX2-2::HSF2 1.00763
X	66543803	66543807	NM_000044	66546894	1	-	HNF-1::TBP::TEF::POU3F2 1.02288
X	1,19E+08	118706185	NM_001000	1,19E+08	-1	-	LHX3::POU3F2::C/EBPalpha::CART-1::CRX 1.07987

Chapter 4

Relapse free survival curves

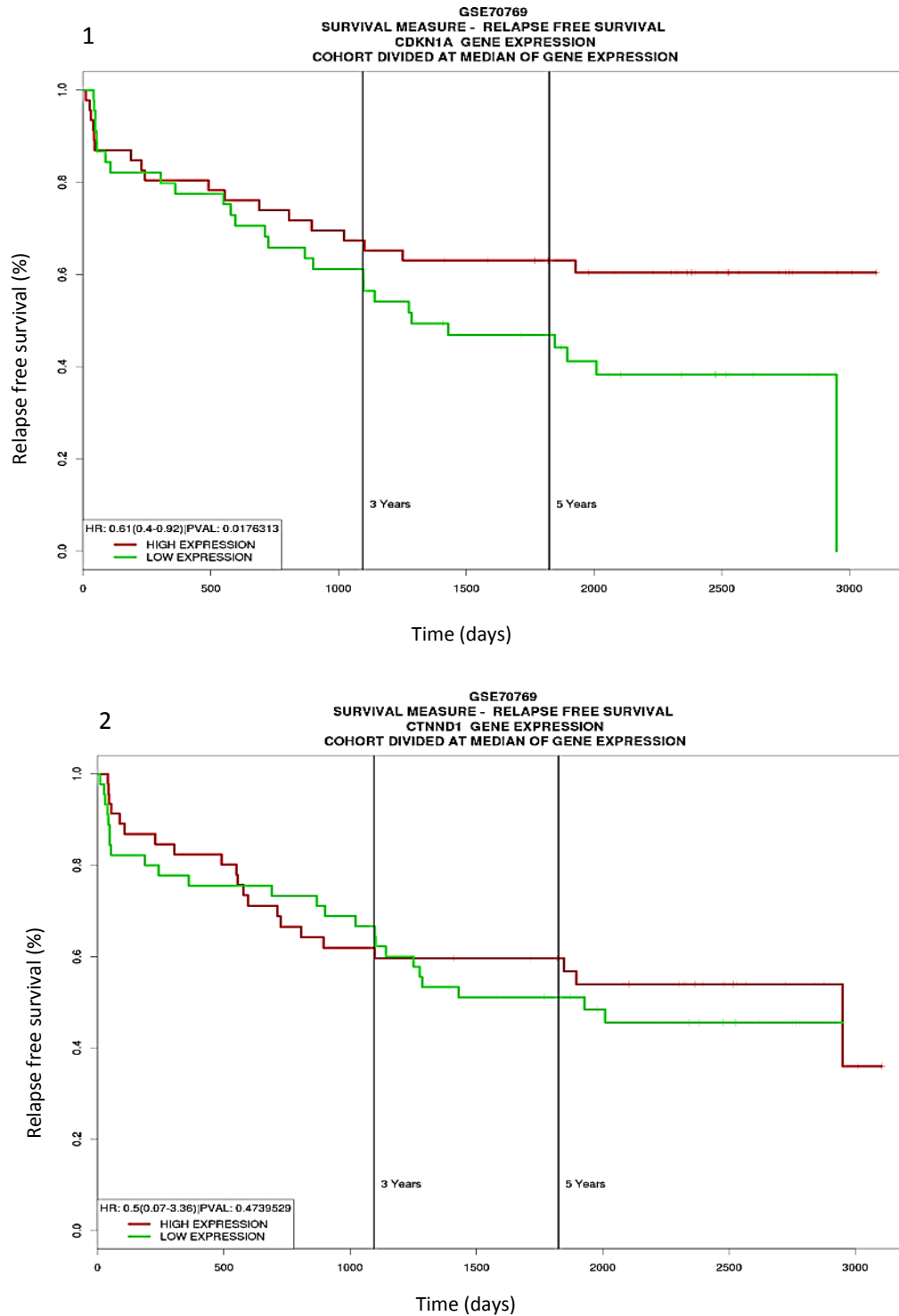


Figure 7.1: Kaplan Meier analysis plots for CDKN1A (1) and CTNND1 (2) for prostate cancer relapse free survival in the ProGene server in the GSE70769 dataset.

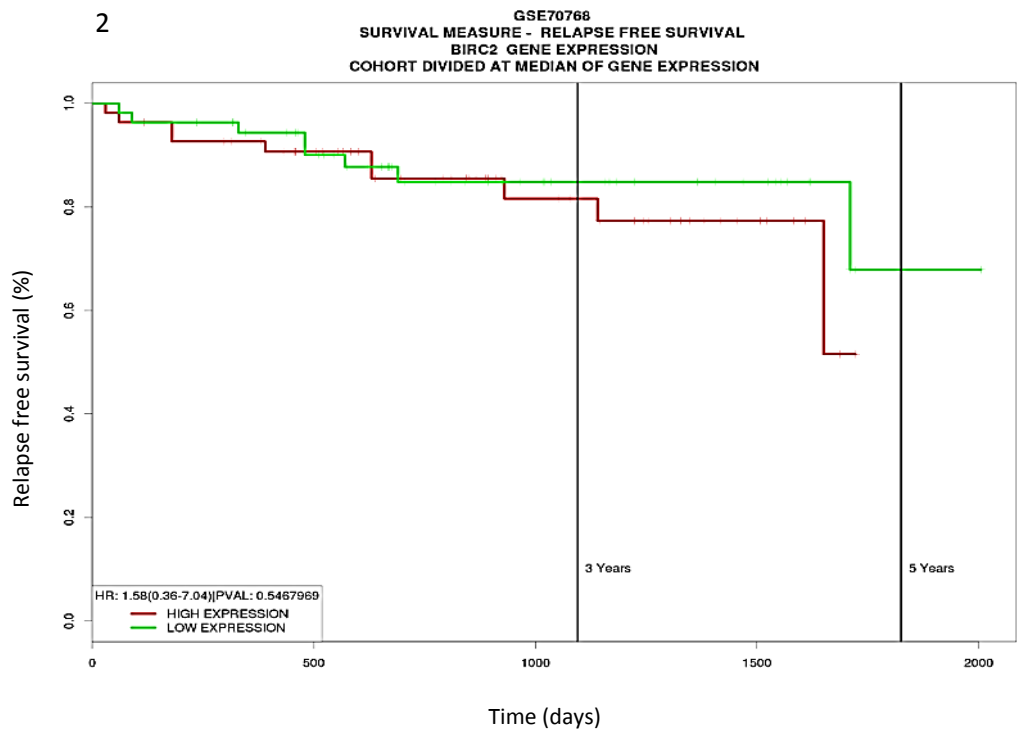
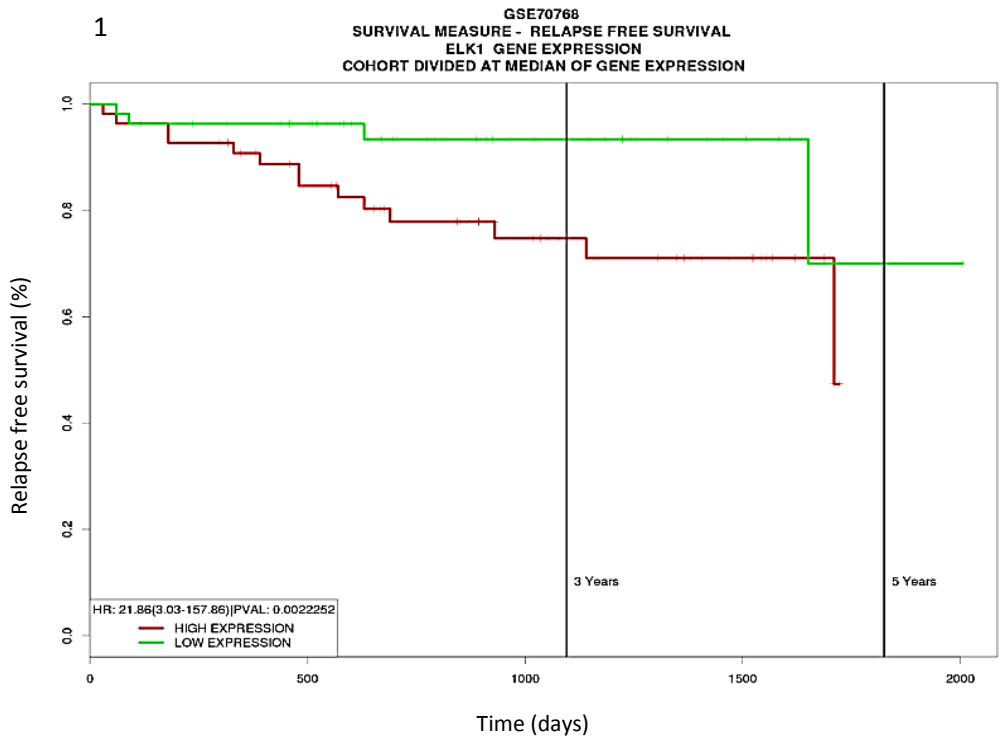
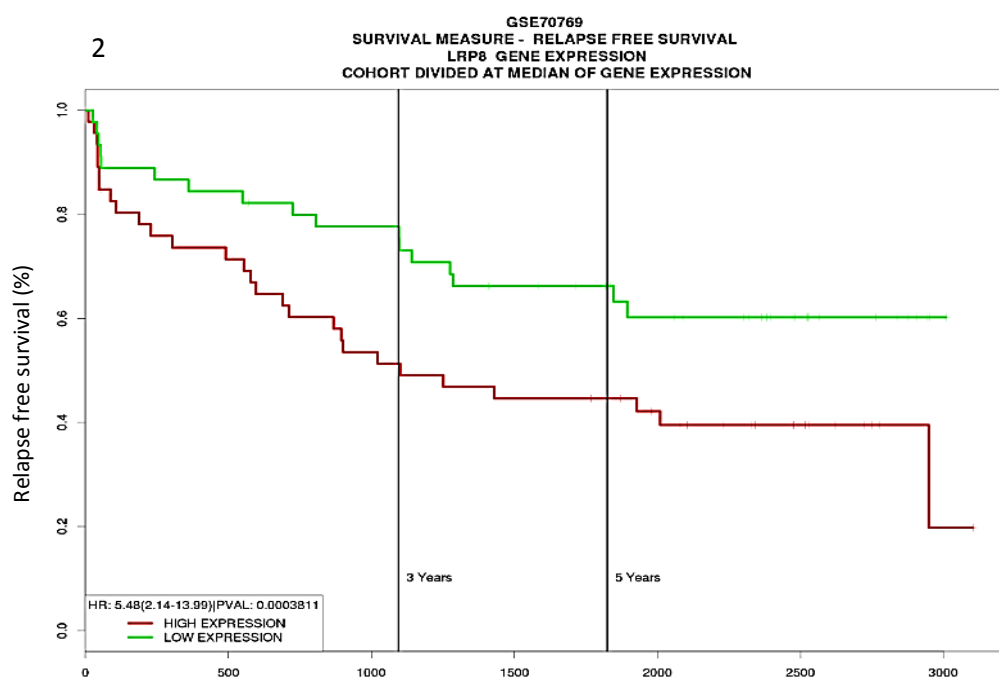
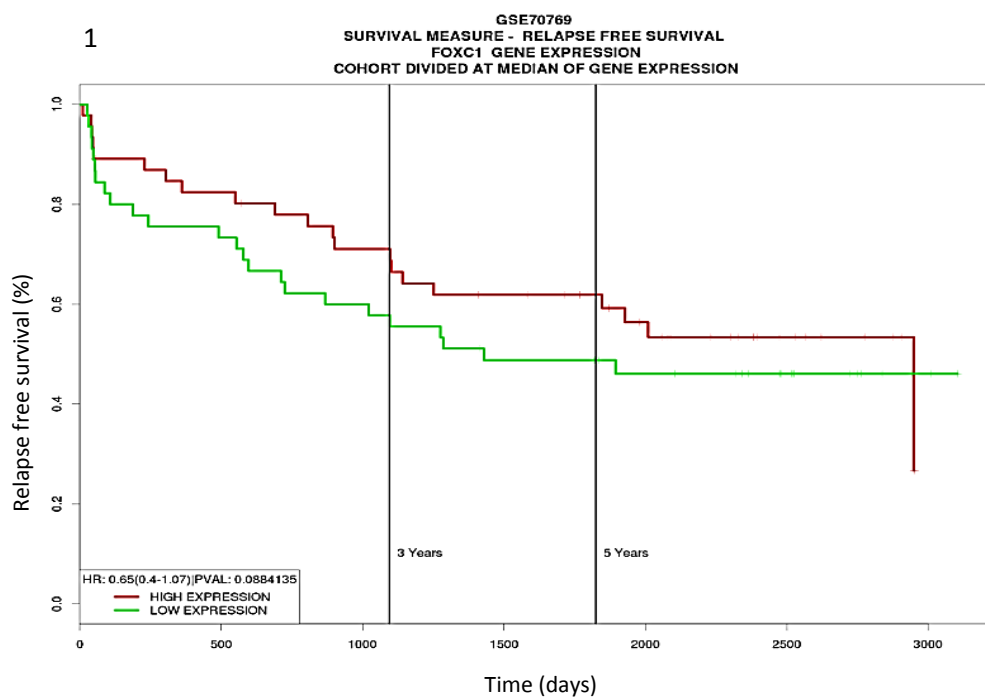


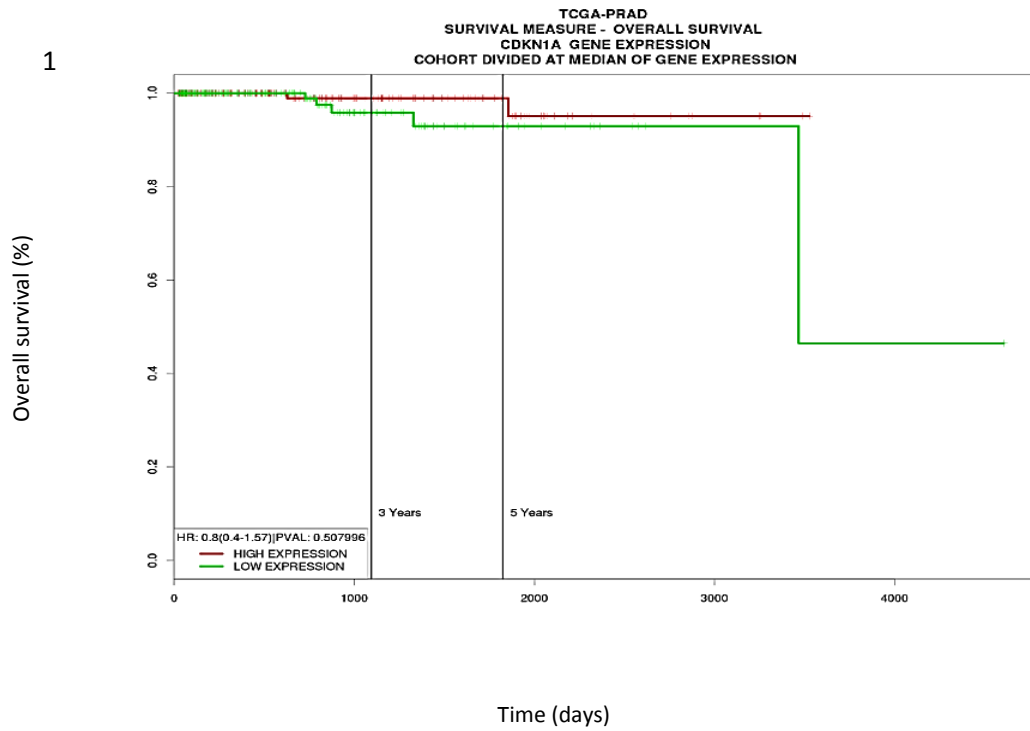
Figure 7.2: Kaplan Meier analysis plots for ELK1 (1) and BIRC2 (2) for prostate cancer relapse free survival in the ProGene server in the GSE70769 dataset.

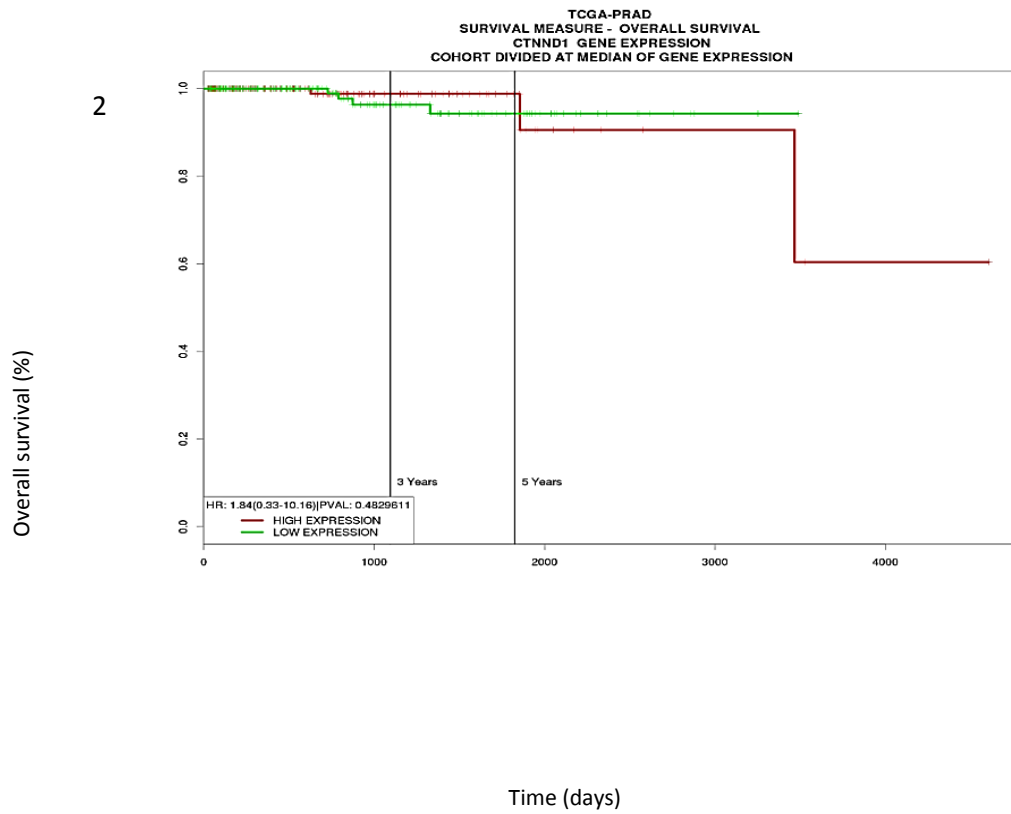


Time (days)

Figure 7.3: Kaplan Meier analysis plots for FOXC1 (1) and LRP8 (2) for prostate cancer relapse free survival in the ProGene server in the GSE70769 dataset.

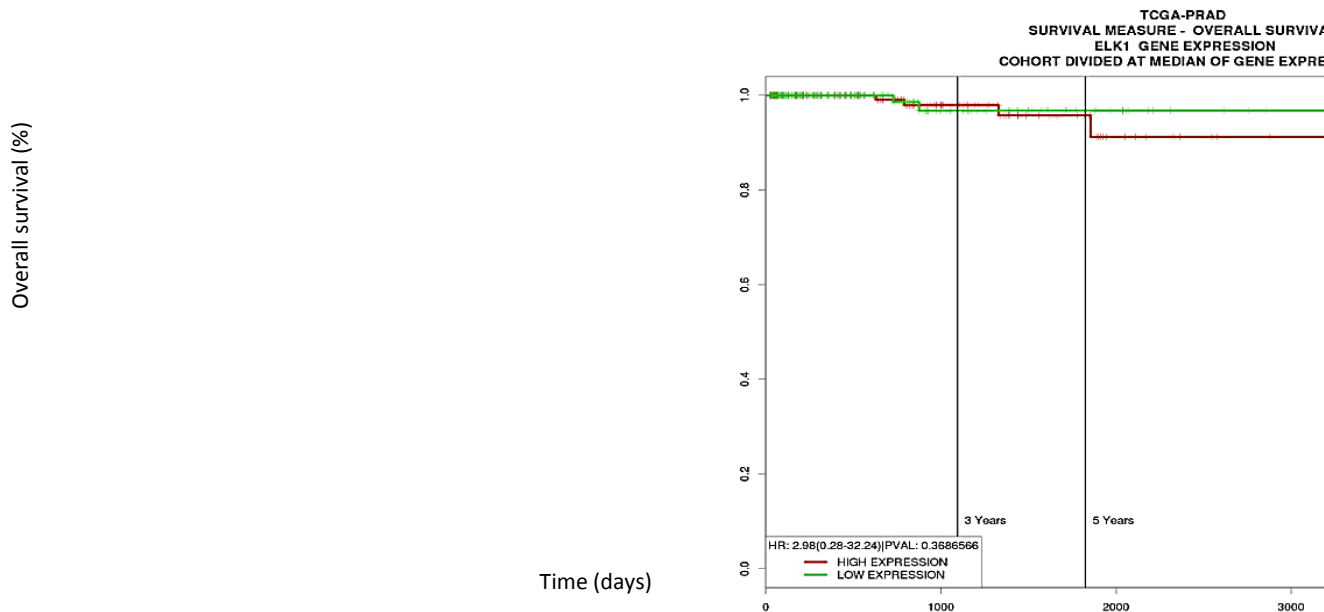
Overall Survival





1

Figure 7.4: Kaplan Meier analysis plots for CDKN1A (1) and CTNND1 (2) for prostate cancer relapse free survival in the ProGgene server in the TCGA-PRAD dataset.



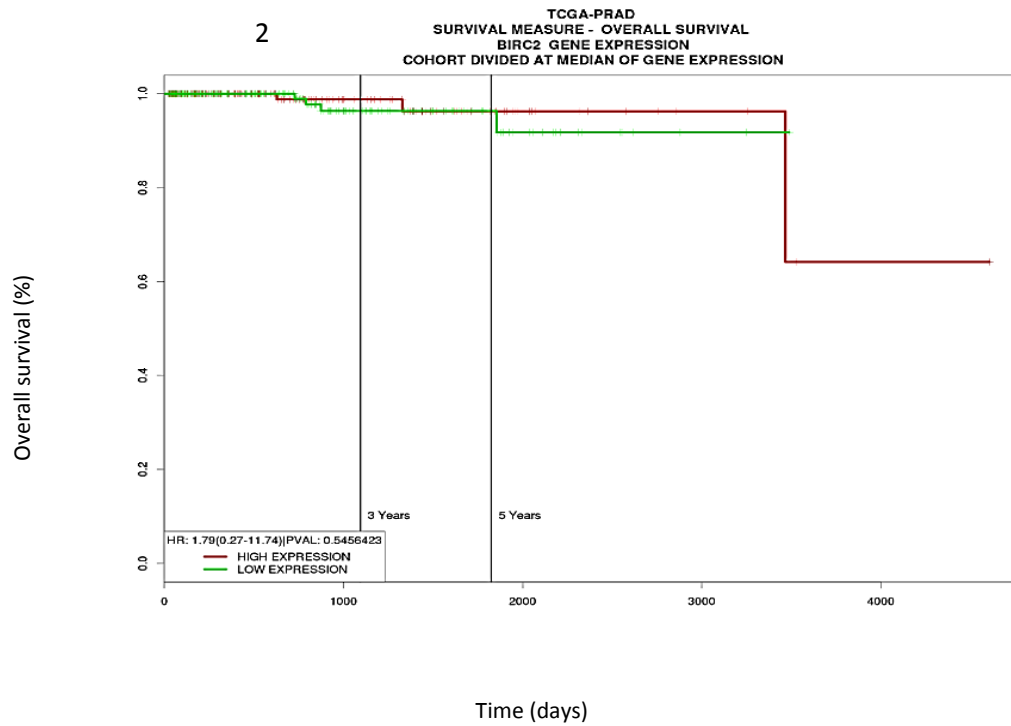


Figure 7.5: Kaplan Meier analysis plots for ELK1 (1) and BIRC2 (2) for prostate cancer relapse free survival in the ProGene server in the TCGA-PRAD dataset.

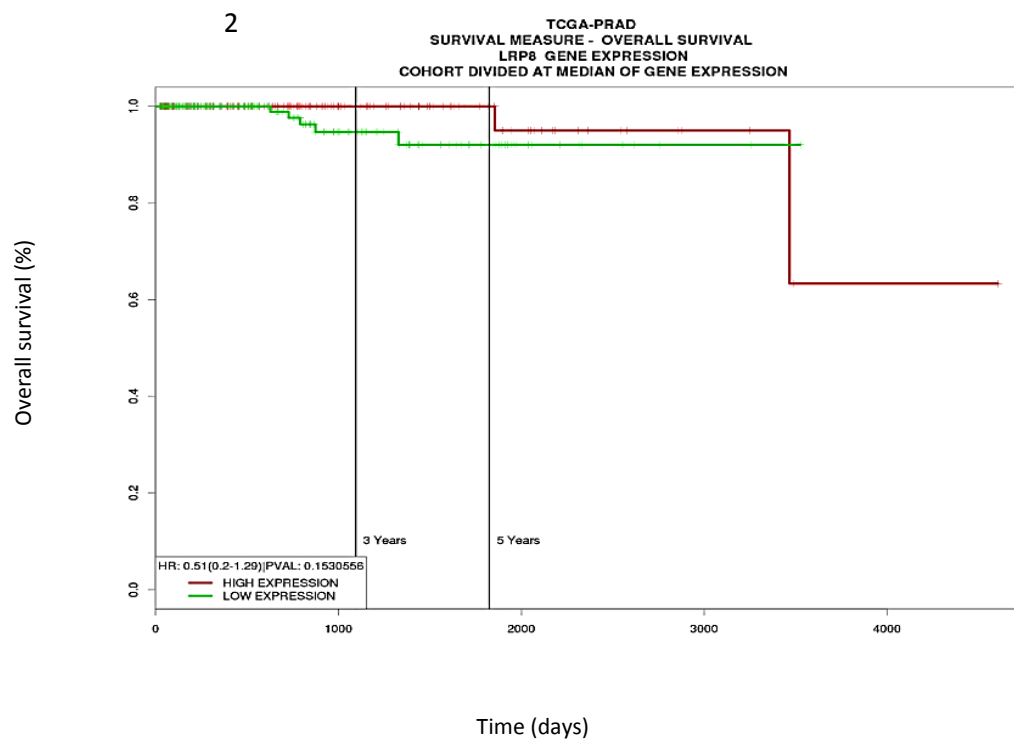
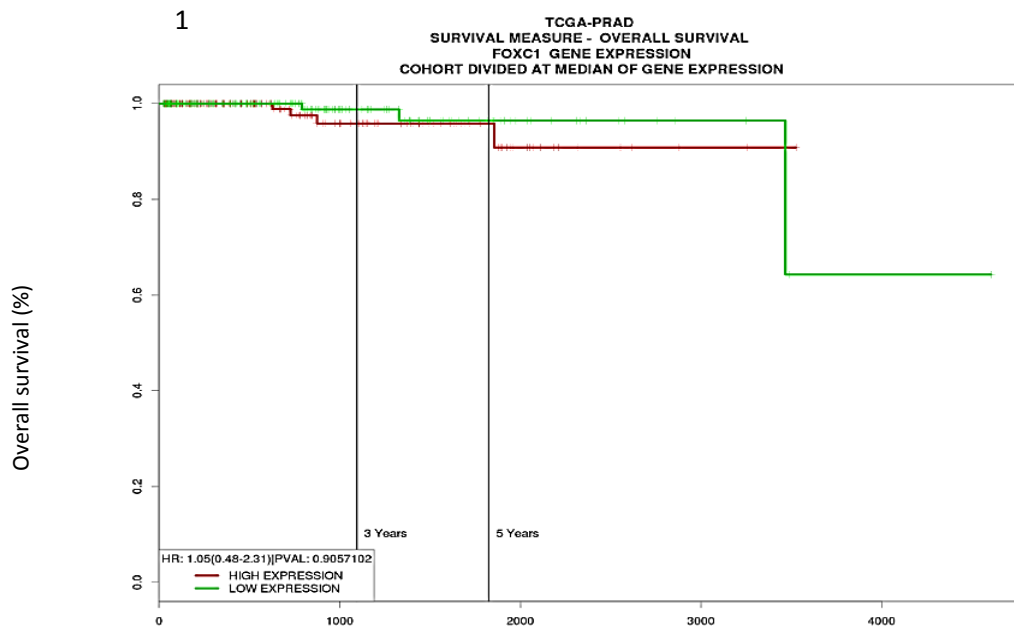


Figure 7.6: Kaplan Meier analysis plots for FOXC1 (1) and LRP8 (2) for prostate cancer relapse free survival in the ProGene server in the TCGA-PRAD dataset.

