# Claremont Colleges Scholarship @ Claremont

KGI Theses and Dissertations

KGI Student Scholarship

7-31-2018

# Computational analysis of genomic variants affecting predicted microRNA:target interactions in prostate cancer.

Angélica Paola Hernández Pérez Keck Graduate Institute, KGI

#### **Recommended** Citation

Hernández Pérez, Angélica Paola. (2018). Computational analysis of genomic variants affecting predicted microRNA: target interactions in prostate cancer. KGI Theses and Dissertations, 6. https://scholarship.claremont.edu/kgi\_\_theses/6. doi: 10.5642/kgitd/6

This Open Access Dissertation is brought to you for free and open access by the KGI Student Scholarship @ Claremont. It has been accepted for inclusion in KGI Theses and Dissertations by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

Computational analysis of genomic variants affecting predicted microRNA:target interactions in prostate cancer

By Angélica Paola Hernández Pérez

A Dissertation submitted to the Faculty of Keck Graduate Institute of Applied Life Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Life Sciences

> Claremont, California 2018

> > Approved by:

Resurdan Antonio

Animesh Ray, PhD Osbaldo Resendis-Antonio, PhD

Copyright by Angélica Paola Hernández Pérez, 2018

All Rights Reserved

We, the undersigned, certify that we have read this dissertation of Angélica Paola Hernández Pérez and approve it as adequate in scope and quality for the degree of Doctor of Philosophy.

Dissertation Committee:

Animesh Ray, Chair

Alexander Zambon, Member

Leander Antonio

Osbaldo Resendis-Antonio, Visiting Examiner (Instituto Nacional de Medicina Genómica)

(for Per

Ranjan Perera, Visiting Examiner (Sanford Burnham Prebys Medical Discovery Institute)



Srikanth Kolluru, PhD Program Director

Abstract of the Dissertation

# Computational analysis of genomic variants affecting predicted microRNA:target interactions in prostate cancer

By Angélica Paola Hernández Pérez

Keck Graduate Institute of Applied Life Sciences: 2018

Prostate cancer (PCa) is the most common cancer of men in the United States and is third only to lung and colon as a cause of cancer death. Clinical behavior of the disease is variable and the combination of prostate-specific antigen (PSA) screening and Gleason score staging are currently the best available molecular and pathology tools to predict outcomes. Cancer biology research establishes microRNAs (miRNAs) as key molecular components in both normal and pathological states. Thus, elucidating miRNAs perturbed by genomic alterations will expand our understanding of the molecular taxonomy of PCa with the aim to complement current practices in the diagnosis, prognosis, and treatment of the disease. This study reports the computational analysis of genomic variants affecting the seed sequence of five miRNAs, changing the prediction of microRNA:target interactions in PC3, an androgen-independent cell line that closely resembles prostatic small cell neuroendocrine carcinoma (SCNC). Genomic variants were detected via deep-sequencing of PC3 and further computational work focused on mapping changes within the seed sequence of predicted mature miRNAs. Five microRNA candidates (from now on denominated microRNA\*) with changes in the g2-g8 seed region were selected: miR-3161\*-5p with rs35834266 G>insA; miR-3620\*-5p with rs2070960 C>T; miR-1178\*-5p with rs7311975 T>G; miR-4804\*-5p with rs266435 C>G; and miR-449c\*-3p with rs35770269 A>T. Subsequently, the computational prediction of miRNA\*:target interactions revealed 643 new relationships. After functional enrichment analysis of new targets, seven genes were associated with endocrine resistance (ABCB11, CDKN1B, NOTCH2, SHC4, CCND1, SP1, ADCY2) and five genes with endocrine and other factor regulated calcium reabsorption (ATP1A2, ESR1, PRRKCB, AP2B1, SLC8A1) categories. A gene-disease association literature search was performed for each of the aforementioned genes in order to understand if they have been implicated in cancer, where CDKN1B, NOTCH2, CCND1 have been reported to participate in prostate cancer progression. Microarray gene expression analyses showed that few predicted microRNA\* targets were underexpressed in untreated PC3 samples versus prostate epithelial cells from the GEO database. However, after assessing the frequency of observed underexpressed genes per candidate microRNA\* using a Fisher's exact test, miR-4804\*-5p target genes (TNKS and GUCY1A3) were statistically significant. Next steps included the comparison between groups of genes subject to non-mutated microRNA and mutated microRNA\* regulation using a Kruskal-Wallis non-parametric test. Results were consistent with the microRNA-gene expression regulation model despite the genomic variant in the seed region, nevertheless the effect of miR-3161\*-5p, miR-3620\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p cannot be predicted solely with the indirect experimental approach that microarray gene expression platforms provide. For this reason, the assessment of recurrent pairwise microRNA-mRNA expression associations was performed using CancerMiner, an online tool from The Cancer Genome Atlas (TCGA) based on a multivariate linear model and rank transformations. Only the relationship of miRNA-3161:CDKN1B was retrieved as a recurrent expression association in uterine corpus endometrial carcinoma (UCEC). In the context of this study, results suggest that CDKN1B (p27<sup>Kip1</sup>) dysregulation by miR-3161\*-5p would be leading to PC3 super proliferation due to the lack of cell cycle arrest from phase G1 to S. Prostate cancer cell line PC3 has shown to share features with prostatic small cell neuroendocrine carcinomas (SCNC) with the implication that molecular mechanisms and therapeutic efficacies observed with PC3 cells are likely applicable to SCNC<sup>1</sup>. Prostatic small cell neuroendocrine carcinoma is a variant form of prostate cancer often characterized by an aggressive course with a poor response to conventional androgen deprivation therapy (ADT), consistent with the lack of the androgen receptor in prostatic small cell carcinoma (SCC)<sup>2</sup>. In some men treated with ADT, development of small cell carcinoma might represent the "escape" of a subpopulation of hormone-independent cells resulting from the selective pressure of hormonal therapy<sup>3</sup>. Hence, the suggestion of CDKN1B dysregulation by miR-3161\*-5p might go beyond the idiosyncrasy of the PC3 cell line, but rather an interesting future direction to investigate prostate cancer patients with SCNC rendering to an adverse disease outcome due to uncontrolled cell proliferation.

# Dedication

To my family and friends, but especially to my mother Laura

for her unconditional love.

Por el aquí y ahora, mañana y siempre.

#### Acknowledgements

I am thankful for the gracious support and encouragement from my advisors, Animesh Ray and Osbaldo Resendis-Antonio, during the course of my graduate work. I thank Alexander Zambon, thesis committee member, for his helpful advice and feedback on the present manuscript, as well as to Ranjan Perera and Bongyong Lee from Sanford Burnham Prebys Medical Discovery Institute since my PhD thesis is based on their experimental work. I am grateful for the graduate students, postdocs and researchers who mentored me as a visiting doctoral student at the Instituto Nacional de Medicina Genómica (México) throughout these years. In particular, to Fabián Flores-Jasso who was instrumental in patiently explaining the Argonaute biochemical assays he developed to elucidate its mechanism of action. I have also had countless discussions with Said Muñoz-Montero and Erick Muciño-Olmos on the computational analysis of microRNA that have broadened the scope of my thinking and heavily impacted the work presented in this thesis. I also thank long-time colleagues from the research community for their support and kindest considerations. Vikram Agarwal was crucial in helping me to understand the target prediction algorithms he developed for TargetScanHuman v7.1 along with the opportunity to explore interesting questions concerning the impact of genomic variants in microRNA. Medical and pathology anatomy colleagues include Angélica Cañedo-González and Bernardo de León Garza who constantly exchanged ideas about the clinical management of prostate cancer and shared valuable literature on their areas of expertise. I thank the Center for Writing & Rhetoric at Claremont Graduate University for providing me the tools and positive feedback to gain confidence on my scientific writing, and to Ana Ruiz-González for her guidance throughout the artwork on this thesis. Finally, I thank all my friends and family for supporting me throughout my life and for inspiring an interest in exploring scientific questions early on. Angélica Paola Hernández Pérez is a doctoral student from the School of Applied Life Sciences, Keck Graduate Institute (KGI) and received fellowship 353554 from CONACyT.

# Contents

Chapter 1 – Organization of the Thesis	1
Context of Study	1
Statement of Problem	2
Aim and Scope	2
Significance of the Study	3
Overview	3
Chapter 2 – Background	5
Clinical Assessment of Prostate Cancer	5
Genomic Variants in Prostate Cancer	7
What is a microRNA?	
MicroRNA, Genomic Variants, and Prostate Cancer	15
Chapter 3 – Research Approach and Methods	18
Cell Lines	19
Custom Target Enrichment Design and Sequence Capture Libraries	20
Deep Sequencing and Variant Calling	21
Precursor microRNA Sequence Retrieval and Mature microRNA Curation	21
Genomic Variant Mapping in microRNA and Candidate Selection	23
Sequence Retrieval of 3'-UTR of mRNAs	24
Target Prediction	25
Functional Enrichment Analysis	26
Gene Expression Analysis	26
Statistical Analysis	27
Recurrent microRNA-mRNA Expression Associations from TCGA	28
Chapter 4 – Results	
Deep Sequencing of pre-microRNA in PC3 and Variant Calling	31
Unique Genomic Variants in microRNA Seed Sequence	32
Target Prediction	34
Functional Enrichment Analysis	35
Gene Expression Analysis	
Statistical Analysis	
Recurrent microRNA-mRNA Expression Associations from TCGA	
Chapter 5 – Discussion and Future Directions	46
References	51
Appendix	62

#### Table of Figures

**Figure 2** Summary of research strategies currently employed by the associated consortia to evaluate the clinical utility of prostate cancer susceptibility loci. In addition, The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC) are the two main consortia providing cancer genomics information related to fifty different cancer types, including prostate cancer (PRAD projects). Image used and modified from C.L. Goh *et al*<sup>62</sup>......8

**Figure 11** Example of precursor hsa-mir-21b and the curation process of its predicted mature microRNA sequences. It represents the integration of both miRBase v.21 and GRCh38/hg18 coordinate systems. The seed sequence region is determined as the first eight nucleotides starting on the 5' end verified by comparing to a list of microRNA families from the

TargetScanHuman v7.1 database. The star represents the location of the genomic variant mapping within the seed sequence of the predicted mature microRNA with canonical-binding....23

**Figure 12** TargetScanHuman v7.1 sample UTR profile for human UBE2D3. The 3'-UTR profiles were constructed using 3P-seq tags (with the number of tags for the 3'-UTR labeled on the y-axis on), which indicate the location and usage of mRNA cleavage and polyadenylation sites. 3P-seq tags from multiple cell lines or tissues were normalized to each other (to account for variable sequencing depth), and then aggregated into one consensus set of counts. Normalized 3P-seq tags were assigned to the representative 3'-UTR of each stop codon and summed (as indicated at the left side of the profile) to provide quantification for the usage of that stop codon. This sum (2312 in the UTR profile shown above) also includes 5 pseudocounts added at the distal end of the Gencode annotation. Image source: http://www.targetscan.org/docs/UTR\_profiles.html .......25

**Figure 18** CDKN1B dysregulation by miR-3161\*-5p might lead to PC3 proliferation due to the lack of cell cycle arrest from phase G1 to S. Molecular mechanisms and therapeutic efficacies observed with PC3 cells are likely applicable to small cell neuroendocrine carcinoma (SCNC), a variant form of prostate cancer often characterized by an aggressive course with a poor response to conventional androgen deprivation therapy (ADT). The pathway from CDKN1B to E2DFP was obtained from KEGG.

#### List of Tables

**Table 2** Seed sequences of normal and mutated microRNA are shown in this table with the corresponding alternative allele insertion reported from PC3 deep-sequencing experiments......33

 Table 4
 Gene enrichment analysis results of genes predicted to be targets of mutated microRNAs. Gene-disease associations include different cancer types where CDKN1B, NOTCH2, and CCND1 have been implicated in prostate cancer progression.
 37

 Table 6 Contingency table miR-4804\*-5p predicted targets. The observed underexpressed genes of miR-4804\*-5p:target interactions are statistically significant.

 40

 Table 10
 Contingency table miR-449c\*-3p predicted targets. The observed underexpressed genes of miR-449c\*-3p:target interactions are not statistically significant.

**Table 11** Kruskal-Wallis rank sum test results. Only the comparison of the medians of all PC3expressed genes inquired by the hgu133plus2 platform (Group A) versus the expressed genes ofuntreated PC3 samples involved in the 643 gained interactions (Group B) and 546 lostinteractions (Group C) were statistically significant.44

# List of Abbreviations and Symbols

PCa	Prostate Cancer	
SCNC	Small Cell Neuroendocrine Carcinoma	
LUTS	Lower Urinary Tract Symptoms	
BPH	Benign Prostatic Hyperplasia	
DRE	Digital Rectal Examination	
PSA	Prostate Specific Antigen	
AR	Androgen Receptor	
ADT	Androgen Deprivation Therapy	
DHT	Dihydrotestosterone	
CRPC	Castration-Resistant Prostate Cancer	
HRPC	Hormone-Refractory Prostate Cancer	
GWAS	Genome-Wide Association Studies	
SNP	Single Nucleotide Polymorphism	
miRNA or miR	microRNA	
microRNA*	Mutated microRNA	
mRNA	messenger RNA	
3'-UTR	3' untranslated region of an mRNA	
pri-miRNA	VA Primary microRNĂ	
pre-miRNA	Precursor microRNA stem loop	
. miRNA	Mature microRNA	
TRBP	Trans-activation-responsive RNA-binding	
	protein	
Drosha	Ribonuclease III. Double-stranded RNA-	
	specific endoribonuclease	
DGCR8	DGCR8 microprocessor complex subunit	
Dicer	Ribonuclease III. Double-stranded RNA-	
	specific endoribonuclease	
XPO5	Exportin-5	
AGO	Argonaute protein	
RISC	RNA-induced silencing complex	
PC3	Androgen-independent prostate cancer cell	
	line PC3	
HPrEpiC or EPIC	Human prostate epithelial cells	
Seed sequence	Nucleotides 2 to 8 within the 5' end of a	
	microRNA that bind predominantly to	
	nucleotides on the 3' of mRNA	
Canonical sites	microRNA binding sites with 8-mer, /-mer At1,	
NOO	and /-mer sequence complementarity	
NGS	Next-generation gene sequencing	
microRinA:target	MicroRina Interacting with mRina target sites	
ININ	Non-mutated microRNA interacting with a non-	
MAN	Mutated migra DNA interacting with a pap	
IVIIN	mutated mRNA target site	
miPRaca	mutated menne target site	
IIIIIIDase	minute and appointer of microniva sequence data	
context++ (TargetScan)	Ruled-based algorithm for target prediction at	
	least as accurate as other existing in vivo	
	crosslinking techniques such as AGO-PAR-	
	CLIP	
dbSNP	The Single Nucleotide Polymorphism	
	J	

	database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms.
AGO-PAR-CLIP	Crosslinking experimental approach to assess microRNA:target interactions
GRCh37/hg19	Human reference genome assembly 37 released on February 2009
GRCh38/hg18	Human reference genome assembly 38 released on December 2013
WebGestalt	Gene enrichment analysis online tool
ORA	Overrepresentation Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes Pathway Database
FDR	False Discovery Rate
BH	Benjamini-Hochberg method
REC	Association-recurrence statistical method
LBC	Luminal Breast Cancer
UCEC	Uterine Corpus Endometrial Carcinoma
SCC	Small Cell Carcinoma

# Chapter 1 – Organization of the Thesis

#### Context of Study

Prostate cancer is the most common cancer among men in the United States and is third only to lung and colon as a cause of cancer death. In 2017, an estimated 161,360 new cases of prostate cancer were newly diagnosed while 26,730 people succumbed to this disease<sup>4</sup>. Multiple genetic and demographic factors, including age, family history, genetic susceptibility, and race, contribute to the high incidence of prostate cancer<sup>4</sup>. Clinical behavior of the disease varies from treating localized tumors with surgery, radiation, and initial hormone therapy to other more aggressive metastasis stages that do not respond to previous hormone treatments<sup>5</sup>. The combination of clinical assessments and pathological parameters such as prostate-specific antigen (PSA) screening and Gleason score staging are currently the best available tools to predict outcomes<sup>6</sup>.

Despite advances in therapy options, the diagnostic landscape has remained relatively static due to poor specificity and sensitivity of screening tools. This results in the over diagnosis of the disease's early-stages and the overtreatment of benign conditions<sup>7</sup>. Hence, there is an urgent need for novel biomarkers to develop alternatives to address these clinical challenges. One of these alternatives are microRNAs, which are small non-coding RNA ~19-25 nucleotides in length that regulate gene expression at the post-transcriptional level<sup>8</sup>. Since the discovery of hundreds of microRNAs in mammals, research on microRNAs has focused on providing evidence in favor of their involvement in normal and diseased conditions. In particular, there has been a community effort to understand their functional implication in the pathogenesis of cancer<sup>9,10</sup>. Thus, elucidating the use of microRNAs as candidate biomarkers can further complement current practices in the diagnosis and prognosis of prostate cancer.

#### Statement of Problem

The emergence of microRNAs in cancer biology has established them as key molecular components in both normal and pathological states<sup>9,11</sup>. The literature on microRNAs has remarkably expanded, with research ranging from assessing the effects of a single microRNA perturbation to understand tumor growth and invasion<sup>12,13</sup>, up to systemic genome-scale reconstructions of cancer hallmarks that include the expression of microRNA signatures<sup>14–16</sup>.

microRNA molecules work as guides to the RNA-induced silencing complex in mRNA destabilization and translational inhibition pathways<sup>17–21</sup>. Once the RNA-induced silencing complex (RISC) is assembled, the Argonaute (Ago) protein binds the mature microRNA which guides the target-recognition process to one or more mRNAs<sup>8,22,23</sup>. microRNAs bind through direct base pairing to potential target sites, prominently in the 3' untranslated region (3'-UTR) of the mRNA<sup>24,25</sup>. A core of 6 nucleotides extending from position 2 to 8 (g2-g8) at the 5' end of the microRNA, termed as the seed sequence, are the most critical for initial target binding<sup>26,27</sup>.

The importance of microRNAs in cancer has been emphasized by the identification of genetic alterations affecting microRNA target sites and the microRNA processing machinery in tumor cells<sup>28,29</sup>. However, it is equally important to understand the effect of genetic variants mapping within the seed sequence of microRNAs. A focus on the gain and loss of microRNA:mRNA interactions due to these variants, especially in the g2-g8 location, is essential to help predicting how a small genetic change can influence target recognition, binding, and the regulation of gene expression profiles in oncogenic events.

#### Aim and Scope

The aim of this study is to computationally predict the loss and gain of microRNA:mRNA interactions due to variants mapping in the seed sequence of microRNAs (miRNA\*) in prostate cancer cell line PC3. The scope of this research is limited to variations in the g2-g8 seed sequence positions associated to functional implications of microRNA fine-tuning of mRNA. This assessment is done in the context of prostate cancer cell line PC3, which closely resembles

prostatic small cell neuroendocrine carcinoma (SCNC), a variant form of prostate cancer that is extremely aggressive and does not respond to hormonal therapy<sup>1,30</sup>.

#### Significance of the Study

The central hypothesis of the present thesis is that a genomic variant mapping within the seed sequence of a microRNA will have an effect on the recognition of microRNA-binding sites thus gaining or losing microRNA: target interactions. If the microRNA: target interaction is lost, the former target gene is expected to be upregulated. If a new microRN:target interaction is gained, the new predicted target gene is expected to be downregulated. This study systematically describes the computational identification of a set of affected microRNAs in the g2-g8 seed sequence region that gain interactions with genes associated with endocrine resistance and reported to be underexpressed in the context of cancer. Results indicate that miR-3161\*-5p (with a rs35834266 G>insA in the seed region) might lead to CDKN1B dysregulation, leading to cell cycle arrest from phase G1 to S thus creating PC3 proliferation. Molecular mechanisms and therapeutic efficacies observed with PC3 cells are likely applicable to small cell neuroendocrine carcinoma (SCNC), a variant form of prostate cancer often characterized by an aggressive course with a poor response to conventional androgen deprivation therapy (ADT). The suggestion of CDKN1B dysregulation by miR-3161\*-5p might go beyond the idiosyncrasy of the PC3 cell line, but rather an interesting future direction to investigate prostate cancer patients with small cell neuroendocrine carcinoma rendering to an adverse disease outcome due to uncontrolled cell proliferation.

#### Overview

This thesis is arranged into five chapters. Chapter 1 contains the introduction encompassing the context, statement of the problem, aim and scope, as well as the significance of the study to guide the reader on a concise, clear and logical structure of the material. Chapter 2 includes the literature background of the clinical assessment of prostate cancer, the current status

of microRNA, and the description of the interplay of microRNA, genomic variants, and prostate cancer. Chapter 3 is a breakdown of the research approach and the description of the computational methods used to fulfill the stated aim of this study. Chapter 4 is the presentation of results. Chapter 5 is the discussion of the study's results and explores possible future research directions. This project has been in collaboration with the Sanford Burnham Prebys Medical Discovery Institute (Lake Nona, Florida), and the Human Systems Biology Laboratory at the Instituto Nacional de Medicina Genómica (CDMX, México).

### Chapter 2 – Background

#### Clinical Assessment of Prostate Cancer

Prostate cancer is the most common cancer among men in the United States and is third only to lung and colon as a cause of cancer death. In 2017, an estimated 161,360 new cases of prostate cancer were newly diagnosed while 26,730 people succumbed to this disease<sup>4</sup>. Prostate cancer mostly affects men in the range of 60 to 80 years old and considering that the worldwide trend is a higher number of aging populations, the disease is predicted to significantly increase in the next decade<sup>31,32</sup>.

Multiple genetic and demographic factors, including age, family history, genetic susceptibility, and race, contribute to the incidence of prostate cancer<sup>4</sup>. Almost 60% of the diagnosed cases are in men 65 years old<sup>33</sup> in average, and the United States the disease is especially frequent in men of African origin, as incidence rates in African Americans are >1.5–fold greater than rates in European Americans<sup>34</sup>. Prostate cancer presents the greatest racial disparity of any cancer in the U.S., and research suggests that environmental influences such as diet and nutrition may have a profound effect on the development and progression of histological cancer to a clinically detectable cancer<sup>35,36</sup>.

The prostate is a walnut-shaped organ that contributes fluid to semen and helps expel semen during ejaculation. Because the prostate lies below the bladder and surrounds the urethra, most prostate cancer signs are tied to urinary disfunction<sup>33</sup>, similar to lower urinary tract symptoms (LUTS)<sup>37</sup>, with a combination of weight loss, bone pain, and lethargy<sup>31</sup>. In general, the earlier the disease is detected, the better the outlook for treatment and arresting cancer progression in patients. Besides the advances, we have to be aware of the risks and adverse effects on the quality of life related to the over-diagnosis and immediate treatment of benign conditions such as prostatitis, inflammation of the prostate and benign prostatic hyperplasia (BPH), a noncancerous enlargement of the prostate gland<sup>38,39</sup>.

Most prostate cancer suspects are identified via digital rectal examination (DRE)<sup>40</sup> (Figure 1) and by raised-levels of prostate specific antigen (PSA). PSA is a glycoprotein secreted by the epithelial cells of the prostate gland, and it is produced to liquefy semen allowing sperm to swim freely. It is expressed in both normal and neoplastic prostate tissues, thus elevated levels of PSA are not specific for malignancy and can be found in conditions such like BPH<sup>41,42</sup>, prostatitis<sup>38</sup>, and urinary tract infections<sup>40</sup>. Approximately 25% of men with PSA levels above 4 ng/mL are diagnosed with prostate cancer, and the risk increases to more than 60% in men with PSA levels above 10 ng/mL<sup>31</sup>. Limitations of PSA screening are well known<sup>43–45</sup> and controversies on its clinical utility are mostly about the "one test, several roles" screening practice. PSA can be used during early detection, risk stratification and staging, as well as in post-treatment monitoring. Thus, recommendations on when and how to use it are beyond our scope and can be reviewed elsewhere<sup>46,47</sup>.



**Figure 1** Digital rectal examination (DRE) is performed by feeling the posterior portion of the gland (peripheral zone) provided that the tumor is sufficiently large to be palpable. Once the tumor metastasizes and migrates to other tissues, biopsies and imaging techniques become resources to locate other tumors with prostate cancer etiology. Image used and modified from Kirby, R. S. and Patel, M. I.<sup>31</sup>

The combination of age-specific PSA, along with pathological findings (Gleason score staging) are currently the best available tools for prognosis and to assess potential treatments for the disease<sup>6,48,49</sup>. Clinical behavior of prostate cancer is variable, with patients having localized tumors that can be treated with radiation therapy (external beam or brachytherapy), several months of hormone therapy, and active surveillance<sup>4,33,46</sup>. The ability to treat many others with

aggressive stages that include bone metastasis and resistance to previous hormone treatments is limited<sup>5,50</sup>.

Androgens and androgen receptor (AR) signaling are necessary for prostate development and homeostasis. AR signaling also drives the growth of nearly all prostate cancer cells. The role of androgens and AR signaling has been well characterized in metastatic prostate cancer, where it has been shown that prostate cancer cells are exquisitely adept at maintaining functional AR signaling to drive cancer growth<sup>51</sup>. Current strategies to restrict its activity are via androgen deprivation therapy (ADT), a standard of care for men with advanced metastatic and recurrent prostate cancer. Dihydrotestosterone (DHT), or  $5\alpha$ -dihydrotestosterone ( $5\alpha$ -DHT), is an endogenous androgen sex steroid and hormone essential for normal prostate growth which also plays a role in the development of the disease<sup>52-54</sup>. ADT acts by blocking DHT and stopping cancer progression with substantial clinical response in conjunction with surgery or radiation<sup>55</sup>. However, ADT becomes less effective over time and does not work in patients with castrationresistant prostate cancer (CRPC)<sup>56</sup>, in which AR is frequently reactivated in the absence of androgens. During intense antiandrogen therapy, a small percentage of men develop treatmentemergent AR-negative small cell/neuroendocrine prostate cancer (SCNC)<sup>57</sup>, a highly aggressive, androgen-independent PCa type that does not respond to ADT at all<sup>58</sup>. Resistance to hormone therapies is a major challenge in PCa creating the need for new prognostic biomarkers and consequently therapeutics targeting signaling pathways that directly or indirectly affect hormone driven gene expression regulation in different levels of the signaling pathways.

#### Genomic Variants in Prostate Cancer

The biological basis of inherited and acquired genetic variants associated with prostate cancer has been studied by several research consortia utilizing genome-wide association studies (GWAS)<sup>34,59</sup>, exome sequencing<sup>60</sup>, and candidate gene studies<sup>61</sup> (Fig 2). Genetic variants that predispose an individual to a particular disease can be attributable to genetic mutations that are rare but highly penetrant, or genetic variants that confer moderate-to-low risk for developing the

disease, or a combination of these factors<sup>62</sup>. Early studies in prostate cancer genetics focused on identifying high-risk loci because these were deemed to have obvious clinical usefulness in terms of identifying a direct causal effect for this disease in individuals, and were easier to identify with existing technologies, like gene expression microarrays<sup>63</sup>.



**Figure 2** Summary of research strategies currently employed by the associated consortia to evaluate the clinical utility of prostate cancer susceptibility loci. In addition, The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC) are the two main consortia providing cancer genomics information related to fifty different cancer types, including prostate cancer (PRAD projects). Image used and modified from C.L. Goh *et al*<sup>62</sup>.

Next-generation sequencing (NGS) offers the advantage to identify variants already annotated and to discover new variants present in the NGS sample of diagnostic interest. Various

published studies using this platform have explored the molecular basis of prostate cancer during the clinical progression of the disease using profiled tumors and model cell lines<sup>63–65</sup>. Reports include multiple recurrent genomic alterations like somatic point mutations, DNA copy-number changes, gene fusions, epigenetic and transcriptomic pathway alterations<sup>66,67</sup>. One NGS tumor sample gives us millions of potentially biologically relevant variables, reminding us that tumor genomics is far more complex and heterogeneous than expected.

The vast majority of our knowledge about cancer genomics comes from the two largest cancer-sequencing consortia, The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC). The aim of both consortia is to completely sequence the genomes of the 50 most important tumor entities. In regards of prostate cancer, The Cancer Genome Atlas reported a comprehensive molecular analysis of 333 primary prostate carcinomas showing seven major genomic subtypes defined by ETS gene fusions (ERG, ETV1/4, and FLI1) or mutations in SPOP, FOXA1, and IDH1 genes. Also, the authors revealed potentially therapeutically actionable lesions in the phosphatidylinositol 3-kinase, mitogen activated protein kinase, and DNA repair genes that are valuable for clinical management<sup>67</sup>. Of high clinical relevance is that each molecular subgroup is defined by a single distinct molecular alteration such as gene fusions or mutations which sets the expectation for a molecular re-classification of prostate cancer.

Despite some evidence of disease-associated coding single nucleotide polymorphisms (SNPs)<sup>68</sup>, the molecular basis of how the majority of the prostate-cancer-associated SNPs function is largely unknown because most are noncoding, lying in intronic or intergenic regions. These SNPs might exert their influence by mapping in regions responsible for fine-tuning regulation of gene expression. Other postulated mechanisms of action include structural rearrangements and changes in DNA structure <sup>23,66</sup>, and RNA structure<sup>69</sup>.

#### What is a microRNA?

microRNAs are small non-coding RNAs ~19-25 nucleotides in length that regulate gene expression at the post-transcriptional level<sup>8</sup>. The biogenesis of animal microRNA is a multistep coordinated process, in which structural and sequential prerequisites are needed for the correct expression of mature microRNA<sup>70</sup>. MicroRNAs are transcribed by RNA polymerase II, resulting in a primary microRNA (pri-miRNA) transcript that is processed by cleavage in the nucleus by a complex involving Drosha and DGCR8 microprocessor complex subunit (DGCR8)<sup>71</sup>. This step forms the precursor microRNA (pre-miRNA) stem loop structure where the 5p strand is present in the forward (5'-3') position and the 3p strand (which will be almost complimentary to the 5p strand) is located in the reverse position.

This pre-miRNA is exported into the cytoplasm by exportin-5<sup>72,73</sup> (XPO5) where it is cleaved by Dicer, the RNase that is found in a multiprotein complex along with the trans-activation-responsive RNA-binding protein (TRBP). In this stage, mature microRNA sequences are formed in a duplex and one strand is incorporated into the RNA-induced silencing complex (RISC)<sup>9</sup> by loading to the Argonaute (AGO1-4) protein (Fig 3). The mature microRNA guides the target-recognition process to one or more mRNAs<sup>8,74</sup>. This step causes the mRNA to be more rapidly degraded or leads to inefficient protein production by reducing protein output<sup>75</sup>.



**Figure 3** Overview of microRNA biogenesis pathway leading to mRNA destabilization and translational repression. In terms of which strand is functional following Dicer cleavage of the stem loop to produce the two mature strands, either the 5p or 3p strand could be functional. The stability of the mature strand may influence its function and ability to enter the RISC complex to then bind to its target gene. In general, the more stable strand will be functional, and the less stable strand will be degraded. In some cases, even both strands could be functional<sup>76</sup>. Image redrawn and modified from Shuibin Lin and Richard I. Gregory<sup>74</sup>.

Understanding the basics of microRNA:mRNA binding mechanisms and its dependence on sequence complementarity is crucial to assess the biological and functional implications in gene expression at the post-transcriptional level. It is well established that microRNAs load on Argonaute proteins<sup>77–80</sup> (in human, AGO2) and are further organized into even smaller segments with thermodynamic<sup>81–83</sup> and kinetic properties<sup>84–90</sup> more typical of RNA-binding proteins than of nucleic acids<sup>27,79</sup>. MicroRNAs bind through direct base pairing to potential target sites prominently in the 3' untranslated region (3'-UTR) of the mRNA. A core of 6 nucleotides extending from position 2 to 8 at the 5' end of the microRNA, termed as the seed sequence, are the most critical for initial target binding<sup>26,27</sup>. Biochemical and computational analyses of preferentially conserved animal microRNAs have revealed how miRNA-pairing motifs interact leading to the identification of several classes of target sites<sup>8</sup>. In the first group, which includes most biologically functional microRNA:target interactions, microRNAs bind to their targets only through their seed regions (the canonical binding). This group can be subdivided into three seed-matching classes: (a) 7mer-A1 site, with sequence complementarity over positions g2-g7 in the guide, plus an A at position 1 on the corresponding sequence on the target mRNA (termed t1); (b) 7mer-m8 site, with sequence complementarity over positions g2-g8; and (c) 8mer site, with sequence complementarity at positions g2–g8 plus an A at t1 (Fig 4).

In the second group, less typical seed matches (the marginal sites) are enclosed: (a) 6mer site, with complementarity at positions g2–g7 but lacking an A at t1; (b) offset 6mer site, with complementarity at positions g3–g8. Whereas most microRNAs bind their targets either canonically or marginally, there exists a third group much less numerous of microRNA:target configuration in which RISC extends the base pairing beyond the seed-match. This type of pairing (atypical sites) either supplements the 6, 7, or 8mer sites described above, or complements for mismatches in the seed region<sup>8,91</sup>.



**Figure 4** microRNA seed sequence categorization and canonical binding sites in the 3'-UTR in mRNA. Most functional microRNAs fall within this categorization. Vertical dashes indicate contiguous Watson–Crick pairing. The probability of conserved targeting and predicted structural accessibility are considered two of the main targeting features in target prediction models such as context++. Image used and modified from David Bartel<sup>8</sup>.

In mammals, Argonaute organizes the guide RNA geometry in a way that positions g2–g7 are in pre-helical form creating a unique configuration that determines whether RISC simply binds, or both binds and cleaves<sup>26,27,87,88</sup>. Argonaute assigns greater value to mismatches before position g5 than from g5–g8. Pairing of g2–g5 with a target appears to be "all or none," with both central and terminal dinucleotide mismatches disrupting binding to a similar extent (Figure 5). Notably, the structure of human AGO2-RISC bound to a seed-matched target suggests that pairing beyond g5 requires a conformational rearrangement in the protein<sup>27,79</sup>.



**Figure 5** Kinetic model for AGO2-RISC function and the impact on target binding. Argonaute assigns greater value to mismatches before position g5 than from g5–g8. TargetScanHuman v7.1 includes predicted seed-pairing stability in the context++ ruled-based algorithm and other fourteen targeting features (Agarwal *et al* 2015<sup>91</sup>) that make this approach at least as predictive as the most *in vivo* cross linking approaches such as PAR-CLIP and HITS-CLIP. Image used and modified from William E. E. Salomon<sup>79</sup>.

Accurate target prediction remains a challenge. Most computational approaches are based on indirect measurements of microRNA and mRNA expression fold changes that are irrelevant to the mechanism of targeting<sup>92–94</sup>. Although numerous advances have been made<sup>95,96</sup>, only TargetScanHuman v7.1 has developed an improved quantitative model (context++) of targeting efficacy using a compendium of 74 experimental datasets pre-processed to minimize confounding biases<sup>91</sup>. The model uses a stepwise regression to identify the most informative features from a large set of potential targeting features<sup>91</sup> (Figure 6). This approach allows context++ to be more predictive than any other published model and at least as predictive as the most informative *in vivo* crosslinking approaches, such as PAR-CLIP and HITS-CLIP<sup>97</sup>.



**Figure 6** Context++ algorithm diagram describing the score features of microRNA, score features of sites, and score features of microRNA families taken into consideration to calculate the context++ score for each microRNA binding site and summary of microRNA:target predictions. Image used from Vikram Agarwal *et*  $al^{\beta_1}$ .

#### MicroRNA, Genomic Variants, and Prostate Cancer

Nearly every microRNA has been associated with abnormal gene expression changes in cancer versus normal cells<sup>98–100</sup>. microRNAs may serve as tumor suppressors, oncogenes, and as housekeeping regulators working as locks or switches when found as signatures<sup>9</sup>. Most of them

are encoded in clusters and work as cooperative functional units with a tendency to target the same gene or different genes in the same pathways<sup>22</sup>. Tissue-specific and stage-specific changes in microRNA targeting are reported to be driven to equilibrate dominant oncogenic signaling pathways active in prostate cancer<sup>97</sup>. Evidence suggests that microRNAs, even in advanced prostate cancer, adapt to regulate continuing alterations to balance oncogenic molecular changes<sup>101–103</sup>.

As any other gene, microRNAs loci are exposed to mutations, amplifications, deletions and genomic translocations. These changes could trigger aberrant expressions of microRNA genes and consequently the dysregulation of target mRNAs<sup>11</sup>. Other indirect causes of dysregulation have been emphasized by the identification of genetic alterations affecting microRNA target sites<sup>23</sup> and the microRNA processing machinery in tumor cells<sup>28,29</sup> such as Drosha<sup>104,105</sup> and Dicer<sup>106</sup>. Understanding genomic alterations in regulatory agents like microRNAs expressed in different stages of prostate cancer are a promising opportunity to add value to current clinical and pathological assessment methods<sup>107,108</sup> (Fig 6).

Since each microRNA generally targets many distinct gene products, genomic variants within the seed of microRNA could produce subtle or drastic changes in the behavior of the cell by perturbing many different genes simultaneously<sup>29,109–111</sup>. The present study identifies, using genome wide targeted deep sequencing, variations within the seed sequence of microRNAs in PC3, an androgen-independent prostate cancer cell line resembling small cell neuroendocrine carcinoma. It provides a novel understanding of the molecular mechanisms behind how signaling pathways related to hormone resistance can be explained by altered microRNA:target interactions due to sequence variations.



**Figure 7** Approaches to study microRNA and their impact in the clinical management of prostate cancer. Evidence suggests that microRNAs dysregulation contributes to prostate cancer initiation and metastatic progression<sup>101–103</sup>. Image used and modified from Josie Hayes<sup>9</sup>

## Chapter 3 – Research Approach and Methods

The goal of this chapter is to introduce the research strategy and methods to perform the computational analysis of genomic variants affecting predicted microRNA:target interactions in the androgen-independent cell line PC3. Genomic variants include single nucleotide polymorphisms, deletions or insertions with at least 1% mutation frequency in the population<sup>112</sup> that have been assigned an rs accession number in the Single Nucleotide Polymorphism database (dbSNP). The central hypothesis of the present thesis is that a genomic variant mapping within the seed sequence of a microRNA will have an effect on the recognition of microRNA-binding sites thus gaining or losing microRNA:target interactions. If the microRNA:target interaction is lost, the former target gene is expected to be upregulated. If a new microRN:target interaction is gained, the new predicted target gene is expected to be downregulated (Figure 7).



Figure 8 Diagram representing the central hypothesis of the present thesis.

To test this hypothesis, prostate cancer cell line PC3 and prostate epithelial cells hPrEpiC were compared to select for unique variants. Sequence capture libraries of precursor microRNAs and genomic variants were designed by Dr. Ranjan Perera and Dr. Bongyong Lee at the Sanford Burnham Prebys Medical Discovery Institute (Lake Nona) in collaboration with NimbleGen<sup>113</sup>. Deep sequencing of PC3 and hPrEpiC sequence capture libraries was performed. Unique genomic variants were filtered computationally focusing on SNPs mapping within the seed

sequence of predicted mature microRNAs. Next steps included the computational prediction of microRNA:target interactions of non-mutated microRNAs (termed microRNA or N) and mutated microRNAs (termed microRNAs (termed microRNAs or M) with potential targets which 3'-UTR mRNA were also sequenced from PC3. The computational prediction of microRNA:target and microRNA\*:target interactions was performed using the rule-based algorithm context++ from TargetScanHuman v7.1. In addition, functional enrichment analysis focused on the genes involved in the gained interactions of microRNA\*:target predictions via WebGestalt<sup>114</sup>. Gene expression microarray data from untreated PC3 and hPrEpiC from the Gene Expression Omnibus (GEO) were retrieved and analyzed using the *affy* package<sup>115</sup> from R Bioconductor. A Fisher's exact test via GraphPad Prism software<sup>116</sup> was used to assess if the observed downregulation of candidate target genes was statistically significant. Finally, the top gained microRNA\*:target interactions derived from this study were queried in CancerMiner, an online repository of statistically recurrent microRNA:mRNA

#### Methods

#### Cell Lines

Prostate cancer cell line PC3 (CRL-1435) was purchased from ATCC<sup>117</sup>. PC3 was derived from a bone metastasis of a grade IV prostatic adenocarcinoma from a 62-year-old Caucasian male. PC3 cells do not express AR and PSA and their proliferation is independent of androgen, similar to small cell neuroendocrine carcinoma (SCNC)<sup>1</sup>. Genomic DNA (gDNA) from PC3 was processed according to ATCC CRL-1435 product sheet<sup>118</sup>. Genomic DNA from human prostate epithelial cells (hPrEpiC) was purchased from ScienCell Research Laboratories<sup>119</sup>. HPrEpiC gDNA was prepared from early passage human prostate epithelial cells using the Qiagen Allprep DNA/RNA Mini Kit<sup>120</sup>. The quality and purity of PC3 and hPrEpiC gDNA was tested by spectrophotometer and gel electrophoresis by Dr. Ranjan Perera and Dr, Bongyong Lee from Sanford Burnham Prebys Medical Discovery Institute (Lake Nona).

#### *Custom Target Enrichment Design and Sequence Capture Libraries*

In collaboration with NimbleGen Corporation, Dr. Ranjan Perera and Dr. Bongyong Lee from Sanford Burnham Prebys Medical Discovery Institute (Lake Nona) developed a custom target enrichment design<sup>113</sup> and sequence capture libraries to identify genome-wide mutations in microRNA and their target binding sites in the 3'-UTR region of mRNAs in PC3. From the entire genome, 1523 microRNA genes annotated in miRBase v.18 were selected to design capture probes of precursor microRNA. For target regions along human UTRs, 67943 non-redundant features of RefSeq genes were selected to design 3'-UTR capture regions. A constant window of nucleotide bases upstream of ATG and a second constant window downstream of the stop codon of each splice variant were downloaded corresponding to each RefSeg sequence. Combined microRNA precursor sequences added up to approximately 124 Kb, and 3'-UTR mRNA regions were approximately 42 Mb. Pairs of unique probe sequences for each identified region were determined using the SSAHA algorithm<sup>121</sup>. Precursor microRNA and 3'-UTR mRNA regions were physically enriched by hybridization following NimbleGen SeqCap Target Enrichment protocols<sup>113</sup>. Nonspecifically bound materials were removed by washing and pull-down materials were amplified and end-repaired for sequencing library preparation and deep sequencing. Figure 9 provides an example of the custom target enrichment design of 3'-UTR mRNA and figure 10 provides an example of the custom target enrichment design of precursor microRNA regions.



**Figure 9** Example of a custom target enrichment design of 3'-UTR mRNA region. 42 Mb Human UTR Design. Capture regions: 3'-UTR. Non-redundant: 67,943 features. Image source: Ranjan Perera and Bongyong Lee.



**Figure 10** Example of a custom target enrichment design of precursor microRNA region. 124 kb microRNA design. Capture regions: microRNA genes. miRBase version 18: 1,523 features. Image source: Ranjan Perera and Bongyong Lee.

#### Deep Sequencing and Variant Calling

The detection of pre-microRNAs and variant calling analysis in PC3 was done by Dr. Ranjan Perera, Dr. Bongyong Lee, and the Sanford Burnham Prebys Medical Discovery Institute Bioinformatics Unit (Lake Nona) using NextGENe v2.3.1<sup>122</sup>. One limitation of this thesis was the lack of provision of BAM and VCF files, adding the need to retrieve precursor sequences from miRBase and to create a purpose-built VCF file by querying dbSNP. Regarding the detection of variants in the 3'-UTR of mRNAs, the same limitation was encountered by the lack of provision of BAM and VCF files. The NextGENe v2.3.1 report of 3'-UTRs included the gene ID associated to one or more transcripts from deep sequencing. These gene IDs were key to retrieve the representative 3'-UTR sequence from the TargetScanHuman v7.1 database. The following sections describe in detail the steps taken to overcome the above technical limitations.

#### Precursor microRNA Sequence Retrieval and Mature microRNA Curation

The microRNA deep sequencing NextGENe v2.3.1 report contained information of precursor miRBase v.18 identification tags and precursor genomic coordinates based on the hum an genome assembly version GRCh37/hg19. The identification tags allowed for the corroboration of annotated precursor sequences in miRBase, however the newest version this database (v.21) is based in GRCh38/hg18. Hence, genomic coordinates had to be converted from GRCh37/hg19

to GRCh38/hg18 using the command-line version of the UCSC Genome Browser LiftOver tool<sup>123</sup>. miRBase is the main online repository of microRNA sequence data and acts as an independent arbiter of microRNA gene nomenclature<sup>124</sup>. Each entry in miRBase v.21<sup>124</sup> represents a predicted hairpin of a microRNA transcript with information on the location of the mature microRNA form. Since each pre-microRNA can have more than one mature form either from the 5' or 3' end of the precursor hairpin, the correct identification of mature sequences is calculated using a coordinates system under miRBase v.21<sup>124</sup> nomenclature.

Take for example hsa-mir-21b. The "hsa" indicates that it is a human microRNA, "mir" refers to its precursor form, and "21" attributes the hairpin to the 21<sup>st</sup> family that was named in the repository and that most likely was discovered earlier. The "b" indicates that it is related to another microRNA probably called hsa-mir-21a. hsa-mir-21b has two mature products, named hsa-miR-21b and hsa-miR-21b\*. The capital R in "miR" indicates the mature sequence form of a microRNA. In this example, miR-21b arises from the 5' arm of the mir-21b hairpin, and miR-21b\* arises from the 3' arm. The (star) "\*" using miRBase nomenclature attributes miR-21b\* as a "minor" product, which means that miR-21b\* can be found in the cell at a lower concentration compared to miR-21b. The final nomenclature for each mature microRNA form would be hsa-miR-21b-5p and hsa-miR-21b-3p, respectively. Now, consider that hsa-mir-21b is located in the chromosome position chr5:195699401-195699497 in regards of the GRCh38/hg18 assembly. However, miRBase reports the location of the ~22 nt long hsa-miR-21b-5p (mature microRNA) ranging from nucleotide 14 to 36 within the pre-microRNA. In order to resolve coordinate systems discrepancies, each pre-microRNA nucleotide has to be counted starting in position 0, to finally get the mature microRNA sequence ranging from nucleotides 14 to 36. This is a computationally exhaustive step which needs to be carefully automated to avoid adding or subtracting nucleotides that will have an impact locating and retrieving the mature microRNA seed sequence. Once genome coordinates for each mature microRNA were located, each sequence was corroborated as a reverse (+) or complementary (-) strand to map the seed sequence. The first eight positions

starting from the 5' end were considered as the seed sequence region, and subsequently verified by comparing to a list of microRNA families from the TargetScanHuman v7.1 database. This reverse-engineering step was crucial for retrieving conserved microRNA family sequences with seed-match canonical binding. Figure 11 describes the process of curating precursor hsa-mir-21b and its corresponding mature microRNA sequences using miRBase coordinates. It represents the integration of both miRBase v.21 and GRCh38/hg18 coordinate systems and already includes how the seed sequence and variant change could be mapped within a predicted mature microRNA.



**Figure 11** Example of precursor hsa-mir-21b and the curation process of its predicted mature microRNA sequences. It represents the integration of both miRBase v.21 and GRCh38/hg18 coordinate systems. The seed sequence region is determined as the first eight nucleotides starting on the 5' end verified by comparing to a list of microRNA families from the TargetScanHuman v7.1 database. The star represents the location of the genomic variant mapping within the seed sequence of the predicted mature microRNA with canonical-binding.

#### Genomic Variant Mapping in microRNA and Candidate Selection

For the purpose of the present study, PC3 was compared to hPrEPiC to identify unique

genomic variants mapping in PC3 microRNA sequences. Changes at the precursor and mature

microRNA sequence level were done by replacing, inserting or deleting the reference nucleotide
to the alternate allele reported by NextGENe v2.3.1. Only genomic variants with an associated Reference SNP (RefSNP) or rs accession number from dbSNP were considered to map statistically significant alleles with a mutation frequency of at least 1% in a population. The definition of variant, mutation, and SNP has generated a lot of discussion and even some contentious disagreement in the research community. However, this is important to understand from the extent of evolutionary conservation. Mutations were originally defined as heritable changes in phenotype and divided either as somatic (non-inherited) or germline (inherited) changes. Variants typically arose as mutations and spread into the population resulting in more than one form termed as "Variant/Alternative form (Allele)". If the mutation frequency is least 1% in a population it can termed as "Polymorphism", mostly Single Nucleotide Polymorphism (SNP). dbSNP is the main information repository of germline and somatic variations in the human species and assigns a RefSNP or rs accession number that appears on maps or graphic representations of the GRCh38/hg18 human genome assembly<sup>112</sup>. In this thesis, each SNP reported by NextGENe v2.3.1 was mapped within the precursor sequence using the chromosome location associated to the rs accession number. If the alternate allele mapped within the region of the potential mature microRNA, then it was verified if it also traced within the g2-g8 region of the seed sequence.

## Sequence Retrieval of 3'-UTR of mRNAs

The verification of 3'-UTR of PC3 mRNAs from deep sequencing was not possible due to the lack of provision of BAM and VCF files. At this stage, two limitations were encountered, 1) the proper identification of delimited 3'-UTR regions from PC3, and 2) the complexity of mapping each potentially affected microRNA binding site without the risk of including false-positives. Using the gene names provided by the NextGENe v2.3.1 report, the representative 3'-UTR sequences of the reported mRNAs were downloaded in FASTA format from the TargetScanHuman v7.1 database. Figure 12 summarizes TargetScanHuman v7.1 annotation process of representative 3'-UTR profiles, and provides an example of a UTR profile for human UBE2D3. The

annotation of each representative 3'-UTR started with the longest Gencode 3'-UTR, extended using information from 3P-seq or other annotation sources<sup>125,126</sup>.

224.2	 ENSTO0000	3343106.5
2012		

TargetScanHuman v7.1 database uses 3'-UTR profiles which represent the prevalence of tandem 3'-UTR isoforms of differing lengths, all of which share the same stop codon<sup>126</sup>. The 3'-UTR profile (red line in Figure 12) drops with each set of clustered 3P-seq tags, indicating the fraction of transcripts that includes that segment of the 3'-UTR. Each TargetScan 3'-UTR profile also shows the location of the most distal end of a Gencode annotation (blue vertical line with Ensembl transcript ID in Figure 12)<sup>91</sup>. It is well known that 3'-UTR regions are shorter in cancerous phenotypes<sup>127,128</sup>, however the decision to use representative 3'-UTR sequences in the present thesis was mainly to create a purpose-built file of verified 3'-UTR sequences for the computational target prediction using the context++ model from TargetScanHuman v7.1.

## Target Prediction

Two computational target predictions were performed using TargetScanHuman v7.1. Since this tool does not have an interface to query a set of mutated microRNAs, the Perl script of the context++ model was downloaded for a custom microRNA:target prediction analysis (code available at: http://www.targetscan.org/cgi-bin/targetscan/data\_download.vert71.cgi). The first prediction was using the set of candidate microRNAs with the alternative allele identified in the seed sequence (from now on termed the mutated microRNA, MN, or miRNA\*:target interactions). The second prediction was using the set of candidate microRNAs without the variant (from now

**Figure 12** TargetScanHuman v7.1 sample UTR profile for human UBE2D3. The 3'-UTR profiles were constructed using 3P-seq tags (with the number of tags for the 3'-UTR labeled on the y-axis on), which indicate the location and usage of mRNA cleavage and polyadenylation sites. 3P-seq tags from multiple cell lines or tissues were normalized to each other (to account for variable sequencing depth), and then aggregated into one consensus set of counts. Normalized 3P-seq tags were assigned to the representative 3'-UTR of each stop codon and summed (as indicated at the left side of the profile) to provide quantification for the usage of that stop codon. This sum (2312 in the UTR profile shown above) also includes 5 pseudocounts added at the distal end of the Gencode annotation. Image source: http://www.targetscan.org/docs/UTR\_profiles.html

on termed the non-mutated microRNA, NN or miRNA:target interactions) using the reference nucleotide from the NextGENe v2.3.1 report. The input files for the microRNA:target prediction using the Perl script were: 1) the microRNA family name, species ID, and the microRNA family seed sequence; 2) the 3'-UTRs sequences of potential targets. For the present thesis, the purpose-built files for the computational target prediction of microRNAs with a genomic variant mapping within the seed sequence included: 1) the microRNA family seed sequences modified with the alternate allele in the g2-g8 positions; and 2) the representative 3'-UTR sequences downloaded from the TargetScanHuman v7.1 database. In order to compare if the target predictions of NN versus MN were different due to the genomic variant, the BEG Venn diagram tool<sup>129</sup> was used to calculate the intersection of the list of elements (MN  $\cap$  NN) and those that were specific to the MN or NN lists. The unique elements of miRNA:target interactions (MN  $\cap$  MN]) were considered lost interactions. The unique elements of miRNA\*:target interactions (MN - [NN  $\cap$  MN]) were considered gained interactions.

### Functional Enrichment Analysis

Functional enrichment of a set of genes was evaluated by the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt)<sup>130,131</sup>, an integrated data mining system for the visualization and statistical analysis of large sets of genes using functional modules within WebGestalt database<sup>114</sup>. A hypergeometric test was performed to assess for an over-representation analysis (ORA) and corrected for multiple hypotheses testing by the Benjamini-Hochberg method<sup>132</sup>. All *p*-values lower than 0.05 were considered as significant. In addition, a gene-disease association literature search was performed of each gene enriched for endocrine resistance and endocrine and other factor-regulated calcium reabsorption categories.

## Gene Expression Analysis

The public Gene Expression Omnibus (GEO) repository was systematically searched for untreated samples of prostate cancer cell line PC3 and for human prostate epithelial cells (hPrEpiC) with the focus on a single DNA microarray platform to obtain easily compared data. For this purpose, the hgu133plus2 human DNA array (Affymetrix Inc., Santa Clara, CA, USA) containing 20,535 DNA probes was selected. Three samples of untreated PC3 (GSM86079, GSM86080, GSM86081) and two samples of human prostate epithelial cells (GSM1095876, GSM1060763) were identified on the hgu133plus2 platform. CEL data files were retrieved from the public GEO repository and processed using a robust multi-array (RMA)<sup>133</sup> normalization protocol with the *affy* package<sup>115</sup> in R Bioconductor. Annotations and attributes were imported from the hgu133plus2 annotation files provided by Affymetrix Inc. (Santa Clara, CA, USA). New predicted target genes from microRNAs\*:target gained interactions were corroborated to have an associated Entrez Gene ID corresponding to any of the 20,535 genes assessed by the hgu133plus2 platform. Overexpressed and underexpressed target genes were assessed using the *limma* package<sup>134</sup>. Differentially expressed genes (DEG) of PC3/hPrEpiC were plotted using a log2 fold change (log2FC) and a *p*-value < 0.01

## Statistical Analysis

A Fisher's exact test was applied to each set of new predicted microRNA\*:target interactions to assess if the observed gene expression changes were statistically significant. Contingency tables were built using Prism v7 (GraphPad Software, La Jolla, CA, USA) and *p*-values lower than 0.05 were considered as significant. In addition, the non-parametric test Kruskal-Wallis was performed using R in order to evaluate if there were statistically significant differences between groups of PC3 expressed genes assessed by the hgu133plus2 platform. These groups were organized as follows: group A included all PC3 genes inquired by the Affymetrix hgu133plus2 microarray minus the total number of genes of gained and lost predicted microRNA:target interactions from mutated and non-mutated microRNAs (total hgu133plus2 genes – genes from the MN list + genes from the NN list). Group B included all lost predicted microRNA\*:target interactions (genes from the MN list). Group D included the predicted microRNA:target interactions (genes from NN list). Group D included the predicted

microRNA:target interactions that remained the same regardless of the genomic variant affecting the seed sequence (genes from the MN  $\cap$  NN lists). Groups A, B, C, D data are available at: https://github.com/microRNA-PCa/PC3/ under the branch names data.A\_BoxPlot, data.B\_BoxPlot, data.C\_BoxPlot, and data.D\_BoxPlot. For the Kruskal-Wallis test, six comparisons were performed using one degree of freedom (df = 1), where *p*-values lower than 0.05 were considered as significant.

## Recurrent microRNA-mRNA Expression Associations from TCGA

Top candidate microRNA\*:target interactions from TargetScanHuman v7.1 which genes were enriched for endocrine resistance and endocrine and other factor-regulated calcium reabsorption from WebGestalt, were queried against the CancerMiner database<sup>14</sup>. CancerMiner is an online resource encompassing a statistical method used to evaluate recurrence of microRNA-mRNA expression association (REC) across cancer types from The Cancer Genome Atlas (TCGA). Figure 13 provides an overview of CancerMiner statistical approach<sup>135</sup>. In individual cancer types, pairwise miRNA-mRNA relationships are evaluated using a multivariate linear model, which also factors in variation (noise) in mRNA expression induced by changes in DNA copy number and promoter methylation at the mRNA gene locus. Associations are rank transformed in individual cancer types, and the method subsequently evaluates the null hypothesis that no association exists between the miRNA-mRNA pair in all cancer types<sup>14</sup>.



**Figure 13** CancerMiner an online tool based on REC, a statistical approach evaluating recurrence of microRNA-mRNA expression association across cancer types from The Cancer Genome Atlas (TCGA). Image used from A. Jacobsen *et al*<sup>14</sup>.

## **Chapter 4 – Results**

#### Summary

In this chapter, the outcomes of the study are thoroughly described starting with variant mapping and microRNA\* selection for further computational target predictions. miR-3161\*-5p, miR-3620\*-5p, miR-1178\*-5p, miR-4804\*-5p, miR-449c\*-3p with a SNP in the g2-g8 position of the seed sequence gained 643 new interactions with genes sequenced from PC3. We provide evidence that alternative alleles mapping within the seed sequence of microRNA do affect the prediction of microRNA\*:target interactions, thus losing or gaining new seed-match canonical binding. After functional enrichment of new targets, ABCB11, CDKN1B, NOTCH2, SHC4, CCND1, SP1, ADCY2 were associated with endocrine resistance. In addition, ATP1A2, ESR1, PRRKCB, AP2B1, SLC8A1 were associated with endocrine and other factor regulated calcium reabsorption. A gene-disease association literature search was performed for each of the aforementioned genes in order to understand if they have implicated in cancer, where CDKN1B, NOTCH2, CCND1 have been associated to prostate cancer progression. Microarray gene expression results showed that few predicted microRNA\* targets were underexpressed in untreated PC3 samples versus prostate epithelial cells from the GEO database. Hence, a Fisher's exact test was performed to assess the frequency of observed underexpressed genes per candidate microRNA\*. Only miR-4804\*-5p was statistically significant with TNKS and GUCY1A3 as target genes. However, neither TNKS, nor GUCY1A3 were part of the group of genes enriched for endocrine resistance. Next steps included comparisons between groups of genes subject to non-mutated microRNA and mutated microRNA\* regulation using a Kruskal-Wallis non-parametric test. The effect of miR-3161\*-5p, miR-3620\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p cannot be predicted solely with this indirect experimental approach. For that reason, the assessment of recurrent pairwise microRNA-mRNA expression associations was performed using CancerMiner, an online tool from The Cancer Genome Atlas (TCGA) based on a multivariate linear model and rank transformations. Results indicate that miR-3161\*-5p and CDKN1B are strong candidates for experimental validation in the context of prostate cancer cell line PC3.

## Deep Sequencing of pre-microRNA in PC3 and Variant Calling

The NextGENe v2.3.1 report of pre-microRNA deep sequencing of PC3 contained the precursor accession number in miRBase v.18, which allowed for the retrieval of precursor sequences from miRBase v.21. In addition, the report included the reference nucleotide (Ref nt), the mutation call (PC3 Mutation Call), and the rs accession number associated to a SNP within dbSNP (SNP db\_xref), allowing for the retrieval of genomic coordinates for each SNP unique to PC3 from the aforementioned database. Table 1 provides a sample of the NextGENe v2.3.1 report and exemplifies genomic variants mapping within precursor microRNAs in PC3 that were not shared with human prostate epithelial cells.

Index	Chr	Mutation	Ref	SNP	PC3	PC3	PC3	PC3	EPIC	EPIC	EPIC	EPIC	Precursor	Precursor
		Chr	nt	db_xref	Covera	Geno	Scor	Muta	Cove	Score	Gen	Muta	miRNA	Accession
		Position			ge	type	е	tion	rage		otyp	tion		
								Call			е	Call		
chr1:228	1	228284991	С	<u>rs2070960</u>	35	CT	12.3	C>C					hsa-mir-	<u>MI0016011</u>
284991								Т					3620	
chr10:10	10	105154089	Α	rs7911488					71	13.1	TC	T>T	hsa-mir-	MI0006444
5154089												С	1307	
chr10:12	10	12695177	Α	rs7896283					130	6.9	GG	A>G	hsa-mir-	MI0016842
695177													4481	
chr10:13	10	135061112	С	rs12355840	21	AA	6.8	G>A	19	6.6	AA	G>A	hsa-mir-	MI0003130
5061112													202	
chr10:29	10	29833998	Α	rs2368393	15	CC	6.1	T>C					hsa-mir-	MI0003617
833998													604	
chr10:29	10	29834003	G	rs2368392	14	TT	6	C>T					hsa-mir-	MI0003617
834003													604	
chr11:48	11	48118350	G	rs35834266	113	insA	15.1	insA	65	6.4	insA	insA	hsa-mir-	MI0014191
118350													3161	
chr11:81	11	81601786	Т	NULL	492	AT	17.9	A>T					hsa-mir-	MI0015831
601786								А					4300	

Table 1 NextGENe v2.3.1 report sample of pre-microRNA deep sequencing and variant calling. A phred-score of QV >30 was considered for all SNP calls.

#### Unique Genomic Variants in microRNA Seed Sequence

The hypothesis of the present thesis is that a genomic variant mapping within the seed sequence of a microRNA will have an effect on the recognition of microRNA-binding sites thus gaining or losing microRNA:target interactions. Unique genomic variants in PC3 were filtered from the NextGENe v2.3.1 report by mapping within precursor sequences using genome coordinates from the GRCh38/hg18. From 52 pre-microRNAs captured in the enrichment design library, 27 genomic variants were unique to PC3 in comparison to human prostate epithelial cells. Only 18 unique variants mapped within predicted mature microRNA sequences, where 12 SNPs mapped within the mature microRNA derived from the 5' end of the reported precursor (-5p) and 6 within reverse/complementary strands derived from the 3' end of reported precursors (-3p).

The molecular mechanism of microRNA:mRNA interaction includes the Argonaute protein, which organizes the guide RNA geometry in a way that positions g2–g7 of the seed sequence end in a pre-helical form. This structural configuration determines whether RISC binds, or binds and cleaves to the target mRNA<sup>27,79,136</sup>. In addition, based on seed-pairing categorization it is known that most microRNA with canonical binding sites in the 3'-UTR in mRNA are functional. With this rationale, only genomic variants mapping within the g2-g8 positions of the seed sequence were considered for the selection of microRNA candidates for target prediction. A total of 5 single nucleotide polymorphisms with an associated RefSNP mapped within the g2-g8 positions of the seed sequence 14 provides a schematic representation of the mapping process of genomic variants from the precursor sequence level to the seed sequence region of predicted mature microRNAs in PC3.



Figure 14 Schematic representation of the process of mapping unique genomic variants of PC3 at the precursor, mature microRNA, and seed sequence sequence level.

The five candidate microRNA sequences were retrieved from miRBase v.21 and verified to have an association to a confidently annotated microRNA family in TargetScanHuman v7.1 database. The reported alternative alleles identified from deep-sequencing were mapped and changed within each mature microRNA candidate sequence as summarized in Table 2.

microRNA	Sequence 5'-3'	dbSNP	Variant call	Seed sequence position
miR-3620-5p	guggg <mark>C</mark> ugggcugggcugggcc	rs2070960	$C\toT$	g6
miR-3161-5p	cugauaagaacagaggcccagau	rs35834266	insA	g8
miR-1178-5p	cagggucAgcugagcaug	rs7311975	$T\toG$	g8
miR-4804-5p	uugga <mark>C</mark> gguaagguuaagcaa	rs266435	$C\toG$	g6
miR-449c-3p	u <mark>U</mark> gcuaguugcacuccucucugu	rs35770269	$A\toT$	g2

**Table 2** Seed sequences of normal and mutated microRNA are shown in this table with the corresponding alternative allele insertion reported from PC3 deep-sequencing experiments.

#### Target Prediction

Two computational target predictions were performed using the Perl script of the context++ algorithm from TargetScanHuman v7.1. Each of the mutated microRNAs was termed with a \* (star) to differentiate the nomenclature between mutated microRNA (microRNA\*) and non-mutated microRNA. The first prediction was using miR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p with the reported alternate allele mapping within the seed sequence from the deep-sequencing analysis as summarized in Table 2. The output file of this first prediction was renamed miRNA\*:target interactions or MN list.

The second target prediction included miR-3620-5p, miR-3161-5p, miR-1178-5p, miR-4804-5p, and miR-449c-3p with the reported reference nucleotide from the deep-sequencing analysis as summarized in Table 2. The output file of the second prediction was termed miRNA:target interactions or NN list. Both computational target predictions were constrained to only assess for canonical binding sites. Finally, both output files were compared (MN list vs NN list) where: 546 interactions were identified as unique elements of the NN list and considered lost interactions; 643 interactions were identified as unique elements of the MN list and considered gained interactions; and 15 interactions were identified as shared elements between the MN and NN lists and were considered as unchanging interactions. Figure 15 provides the Venn diagram of the comparison of the predicted microRNA:target interactions (NN elements) versus microRNA\*:target interactions (MN elements). TargetScanHuman v7.1 predictions results summarized as MN and NN lists are available at https://github.com/microRNA-PCa/PC3



**Figure 15** Venn diagram summarizing the 546 lost interactions (NN), 643 gained interactions (MN), and 15 interactions that were still predicted for both mutated and non-mutated microRNAs (MN  $\cap$  NN).

#### Functional Enrichment Analysis

A functional enrichment analysis was performed on gained target genes (from predicted microRNA\*:target interactions) and lost target genes (from predicted microRNA:target interactions). Both gene sets were queried for an over-representation analysis (ORA) using WebGestalt. Inherent limitations of the ORA method include the mandatory selection of a threshold in the analysis of gene subsets and the statistical assumption that all genes are equal, which might result in loss of information. From 546 predicted genes interacting with non-mutated microRNA, only 510 had an associated Entrez gene ID which was used by WebGestalt for the ORA analysis (Supplementary Table 1). From 643 predicted genes interacting with mutated microRNA\*, only 598 had an associated Entrez gene ID which was used by WebGestalt for the ORA analysis. As multiple tests were carried out simultaneously, a gene was considered to be over-represented if the *p*-value was < 0.05 after the Benjamini-Hochberg adjustment. No enrichment was found for miR-449c\*-3p gained targets, probably due to the small number of candidate targets available from the computational prediction of miR-449c\*-3p:target interactions.

Table 3 summarizes the ORA results of enriched genes from gained microRNA\*:target interactions.

ID	Name	#Gene
<u>hsa01040</u>	Biosynthesis of unsaturated fatty acids - Homo sapiens (human)	4
<u>hsa04977</u>	Vitamin digestion and absorption - Homo sapiens (human)	4
hsa04976	Bile secretion - Homo sapiens (human)	7
<u>hsa05412</u>	Arrhythmogenic right ventricular cardiomyopathy (ARVC) - Homo sapiens (human)	7
<u>hsa01522</u>	Endocrine resistance - Homo sapiens (human)	8
<u>hsa04961</u>	Endocrine and other factor-regulated calcium reabsorption - Homo sapiens (human)	5
<u>hsa04924</u>	Renin secretion - Homo sapiens (human)	6
<u>hsa05224</u>	Breast cancer - Homo sapiens (human)	10
<u>hsa04978</u>	Mineral absorption - Homo sapiens (human)	5
<u>hsa05414</u>	Dilated cardiomyopathy - Homo sapiens (human)	7

**Table 3** Summary of the functional enrichment analysis of gained target genes. 598 Entrez gene IDs out 643 predicted genes were enriched using a hypergeometric analysis. These set of genes have been predicted to interact with miR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, and miR-4804\*-5p due to an alternate allele in the seed sequence region. The first column ID corresponds to the KEGG pathway ID annotated in the human species (hsa), the second column corresponds to the enrichment category, and the third column corresponds to the number of genes that have been associated to each enrichment category.

Seven genes were enriched for endocrine resistance with a *p*-value < 1.28e-02: ABCB11, CDKN1B, NOTCH2, SHC4, CCND1, SP1, ADCY2. And five genes were enriched for endocrine and other factor-regulated calcium reabsorption with a *p*-value < 1.64e-02: ATP1A2, ESR1, PRRKCB, AP2B1, SLC8A1. A genes-disease association literature search was performed for each of the aforementioned genes in order to identify which genes have been implicated in cancer. From the literature search CDKN1B, NOTCH2, CCND1 were identified to participate in prostate cancer progression. Table 4 summarizes the gene enrichment analysis results organized by microRNA\*:target and subdivided by enrichment category. In addition, a brief description of the gene-disease association is included with its corresponding literature reference.

microRNA	Enrichment category	Predicted targets	Disease Association	Ref
	Endocrine resistance	ABCB11	Liver diseases such as cholestasis	137,138
miR-3161*-5p		CDKN1B	Multiple endocrine neoplasia, prostate cancer, breast cancer	139–142
		NOTCH2	Angiogenesis and metastasis, prostate cancer	143–145
		SHC4	Urethral structure and male reproductive organ cancer	146,147
	Endocrine and other factor-regulated calcium	ATP1A2	Migraine and familial hemiplegic	148–150
	reabsorption	ESR1	Breast, endometrial, and ovarian cancer	67,151–153
miR-3620*-5p	Endocrine resistance	CCND1	Multiple myeloma, breast and prostate cancer	142,154–156
		SP1	Huntington disease, gastric, breast, and colorectal cancer	157–160
miR-1178*-5p	Endocrine resistance	ADCY2	Adrenal gland disease and colorectal cancer	161,162
	Endocrine and other factor-regulated calcium reabsorption	PRKCB	Glioblastoma	163,164
miR-4804*-5p	Endocrine and other factor-regulated calcium	AP2B1	Ataxia telangiectasia, rhabdomyosarcoma	165–167
	reabsorption	SLC8A1	Long qt syndrome, heart diseases	168,169

**Table 4** Gene enrichment analysis results of genes predicted to be targets of mutated microRNAs. Genedisease associations include different cancer types where CDKN1B, NOTCH2, and CCND1 have been implicated in prostate cancer progression.

## Gene Expression Analysis

To further assess the hypothesis of the present thesis, if genomic variants significantly affect the seed sequence of microRNA\*, then the expression of genes associated to endocrine resistance should be perturbed in PC3 versus prostate epithelial cells (EPIC). In other words, gained targets are anticipated to be downregulated given the microRNA\*-mediated model for gene expression regulation at the post-transcriptional level. Microarray gene expression analyses of predicted new targets were performed using three untreated PC3 samples and 2 human prostate epithelial cells samples from the GEO database (see Chapter 3 Methods for accession number). Differentially expressed genes (DEG) were assessed by comparing prostate cancer cell

line PC3 versus human prostate epithelial cells (hPrEpiC or EPIC). Gene expression analysis results are displayed in figure 16 using a volcano plot with a *p*-value < 0.01 and a log2 Fold Change. The color coding indicates the overexpression of 374 genes (red) and the underexpression of 152 genes (green) in prostate cancer cell line PC3. The list of overexpressed genes is available at https://github.com/microRNA-PCa/PC3/blob/master/DEG\_over and the list of underexpressed genes is available at https://github.com/microRNA-PCa/PC3/blob/master/DEG\_over.



PC3-EPIC

Figure 16 Differentially expressed genes (DEG) from PC3-EPIC microarray data using the Affymetrix hgu133plus2 platform, where 374 genes were overexpressed (red) and 152 genes were underexpressed (green).

From 643 predicted microRNA\*:target interactions, a total of 592 Entrez gene IDs were assessed by the hgu133plus2 microarray platform. From those 592 Entrez gene IDs, only few genes had expression changes according to the microRNA\*:target model. Table 5 provides a summary of predicted new targets from TargetScanHuman v7.1 by microRNA\* and the number of observed underexpressed genes from those predicted interactions.

Mutated microRNA*	TargetScanHuman v7.1 (predicted gained interactions)	Affymetrix hgu133plus2 platform (underexpressed genes of predicted gained interactions)
miR-3161*-5p	220	2
miR-3620*-5p	90	2
miR-1178*-50	122	3
miR-4804*-5p	156	4
miR-449c*-3p	55	1
Total new interactions	643	12

**Table 5** Summary of predicted new targets by TargetScanHuman v7.1 and the number of observed underexpressed genes from those predicted interactions in the DEG Affymetrix hgu133plus2 data.

The small amount of observed changes was most likely due to the low number of untreated PC3 and human prostate epithelial cell samples in the GEO database. A Fisher's exact test was needed to translate the frequency of perturbed targets into a *p*-value to assess for statistically significance.

## Statistical Analysis

A Fisher's exact test was used to assess if the observed gene expression changes of gained targets were statistically significant (p-value < 0.05). For each microRNA\* a contingency table was built using the Prism v7 software<sup>116</sup> with the values provided in table 5. For example, the contingency table for miR-4804\*-5p has 156 predicted interactions from TargetScanHuman (column A) where only 4 genes are observed to be underexpressed assuming that they are downregulated due to miR-4804\*-5p (column B). The hgu133plus2 platform inquires a total of 20,535 genes (column A) where 152 genes are observed to be underexpressed in PC3 (column B). The Fisher's exact test p-value was 0.0324, meaning that the observed underexpressed genes of miR-4804\*-5p:target interactions are statistically significant (Table 6).

Data analyzed	Column A	Column B	Total
Containing binding sites	156	4	160
All genes in microarray	20535	152	20687
Total	20691	156	20847
<i>p</i> -value	0.0324		
One- or two-sided	Two-sided		
Statistically significant ( <i>p</i> < 0.05)?	Yes		

 Table 6 Contingency table miR-4804\*-5p predicted targets. The observed underexpressed genes of miR-4804\*-5p:target interactions are statistically significant.

Contingency tables were built for each microRNA\* using the Prism v7 software and results were as follows:

Table 7 recapitulates the Fisher's exact test results for miR-3161\*-5p with a *p*-value of 0.6802, meaning that the observed underexpressed genes of miR-3161\*-5p:target interactions are not statistically significant.

Data analyzed	Column A	Column B	Total
Containing binding sites	220	2	222
All genes in microarray	20535	152	20687
Total	20755	154	20909
<i>p</i> -value	0.6802		
One- or two-sided	Two-sided		
Statistically significant ( <i>p</i> < 0.05)?	No		

 Table 7 Contingency table miR-3161\*-5p predicted targets. The observed underexpressed genes of miR-3161\*-5p:target interactions are not statistically significant.

Table 8 recapitulates the Fisher's exact test results for miR-3620\*-5p with a *p*-value of 0.1489, meaning that the observed underexpressed genes of miR-3620\*-5p:target interactions are not statistically significant.

Data analyzed	Column A	Column B	Total
Containing binding sites	90	2	92
All genes in microarray	20535	152	20687
Total	20625	154	20779
<i>p</i> -value	0.1489		
One- or two-sided	Two-sided		
Statistically significant $(p < 0.05)$ ?	No		

 Table 8 Contingency table miR-3620\*-5p predicted targets. The observed underexpressed genes of miR-3620\*-5p:target interactions are not statistically significant.

Table 9 recapitulates the Fisher's exact test results for miR-1178\*-5p with a *p*-value of 0.0668, meaning that the observed underexpressed genes of miR-1178\*-5p:target interactions are not statistically significant.

Data analyzed	Column A	Column B	Total
Containing binding sites	122	3	125
All genes in microarray	20535	152	20687
Total	20657	155	20812
<i>p</i> -value	0.0668		
One- or two-sided	Two-sided		
Statistically significant $(p < 0.05)$ ?	No		

 Table 9 Contingency table miR-1178\*-5p predicted targets. The observed underexpressed genes of miR-1178\*-5p:target interactions are not statistically significant.

Table 10 recapitulates the Fisher's exact test for miR-449c\*-3p with a *p*-value 0.3398, meaning that the observed underexpressed genes from miR-449c\*-3p:target interactions are not statistically significant.

Data analyzed	Column A	Column B	Total
Containing binding sites	55	1	56
All genes in microarray	20535	152	20687
Total	20590	153	20743
<i>p</i> -value	0.3398		
One- or two-sided	Two-sided		
Statistically significant ( <i>p</i> < 0.05)?	No		

**Table 10** Contingency table miR-449c\*-3p predicted targets. The observed underexpressed genes of miR-449c\*-3p:target interactions are not statistically significant.

Finally, only the observed underexpressed genes of miR-4804\*-5p:target interactions were statistically significant. The genes predicted to be interacting with miR-4804\*-5p and underexpressed in PC3 were: TNKS, a poly (ADP-ribose) polymerase involved in the regulation of Wnt/beta-catenin signaling and that has been implicated in the pathology of various forms of cancer<sup>170</sup>. And GUCY1A3, a paralog gene of GUCY1A1, which is a soluble guanylate cyclase that catalyzes the conversion of GTP to 3',5'-cyclic GMP and pyrophosphate, and that has been implicated in progressive cerebral angiopathy<sup>171</sup>. However, neither TNKS, nor GUCY1A3 were enriched for endocrine resistance via WebGestalt.

In addition, two box plots were built using log2 ratios and *p*-values to compare the expression of all PC3 genes assessed by the hgu133plus2 platform versus the expressed genes of untreated PC3 samples involved in the 643 gained interactions, 546 lost interactions, and 15 remaining interactions despite the genomic variant affecting the 5 candidate microRNAs\* seed sequence. Figure 17 summarizes the aforementioned groups, which were organized as follows: group A included all genes inquired by the hgu133plus2 microarray minus the total number of genes of gained and lost predicted microRNA:target interactions from mutated and non-mutated microRNAs (total hgu133plus2 genes – genes from the MN list + genes from the NN list). Group B included all gained predicted microRNA\*:target interactions (genes from the MN list). Group D included the predicted microRNA:target interactions (genes from NN list). Group D included the predicted microRNA:target interactions (genes from NN list). Group D included the predicted microRNA:target interactions (genes from NN list). It was not apparent from the box plots that the average expression levels of gained targets were lower than the total average expression, and this could not be deduced either by comparing different quartile distributions as shown in the box plots.



**Figure 17** Gene expression analysis using box plots to compare the expression of all PC3 genes assessed by the hgu133plus2 platform (Group A) versus the expressed genes of untreated PC3 samples involved in the 643 gained interactions (Group B), 546 lost interactions (Group C), and 15 remaining interactions (Group D) despite the genomic variant affecting the 5 candidate microRNAs\* seed sequence. It was not apparent from the box plots that the average expression levels of gained targets were lower than the total average PC3 gene expression.

Hence, the Kruskal-Wallis non-parametric test was used to determine if there were statistically significant differences between groups. Table 11 summarizes the Kruskal-Wallis test results, where only the comparison of Group A vs Group B, and the comparison of Group A vs Group C were statistically significant. In other words, Group A is not subject to the present study five candidate microRNAs regulation, while Group B is subject to miR-3620-5p, miR-3161-5p, miR-1178-5p, miR-4804-5p, and miR-449c-3p and Group C is subject to miR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p. These results are consistent with the microRNA-gene expression regulation model, despite the genomic variant in the seed region.

Comparison	Kruskal-Wallis chi-squared ( <i>H</i> )	<i>p</i> -value by group	Statistically significant ( <i>p</i> < 0.05)?
Group A vs Group B	60.915	5.959e-15	Yes
Group A vs Group C	30.332	3.641e-08	Yes
Group A vs Group D	3.8363	0.05015	No
Group B vs Group D	0.80014	0.3711	No
Group B vs Group C	1.9647	0.161	No
Group C vs Group D	1.5744	0.2096	No

Degrees of freedom (df = 1)

**Table 11** Kruskal-Wallis test results. Only the comparison of the medians of all PC3 expressed genes inquired by the hgu133plus2 platform (Group A) versus the expressed genes of untreated PC3 samples involved in the 643 gained interactions (Group B) and 546 lost interactions (Group C) were statistically significant.

## Recurrent microRNA-mRNA Expression Associations from TCGA

The small number of untreated PC3 and prostate epithelial cell samples available at the GEO database, and the limitation of not having pairwise microRNA-mRNA expression data in the same physiological conditions made it difficult to garner meaningful information to support functional implications of computationally predicted microRNA\*:target interactions. For that reason, miR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p and the alternate alleles from PC3 deep-sequencing reports were queried using The Cancer Genome Atlas (TCGA) Data Portal interphase. None of the candidate microRNAs or genomic variants retrieved any results, probably because these microRNAs have been newly annotated based on their miRBase nomenclature, their low abundance in different profiled tumor types at TCGA, or lost information due to the lack of uniformity in data generation in TCGA<sup>172</sup>.

However, Jacobsen *et al.* developed a statistical method to evaluate recurrence of miRNAmRNA expression association across ten epithelial cancer types from TCGA. This method considers pairwise miRNA-mRNA relationships in individual cancer types and evaluates these relationships using a multivariate linear model, which also factors in variation (noise) in mRNA expression induced by changes in DNA copy number and promoter methylation at the mRNA gene locus. Associations are rank transformed in individual cancer types, and the method subsequently evaluates the null hypothesis that no association exists between the miRNA-mRNA pair in all cancer types<sup>14</sup>.

The twelve candidate microRNA\*:target interactions summarized in Table 4, were queried against the CancerMiner database. Only the relationship of miRNA-3161:CDKN1B was retrieved as a recurrent expression association in uterine corpus endometrial carcinoma (UCEC) with an association score (REC) by individual cancer type of -0.44 and a *p*-value of 0.366. Although the REC score did not provide a very strong association at the individual level, the result supports that miRNA-3161:CDKN1B interaction is likely to be functionally relevant in the cancer context.

## **Chapter 5 – Discussion and Future Directions**

The present thesis systematically analyzes microRNA seed polymorphisms in prostate cancer cell line PC3 and infers possible effects on microRNA:target interactions with functional implications relevant in the cancer context. MiR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p with a SNP mapping in the g2-g8 seed sequence region gained 643 computationally predicted microRNA\*:target interactions with canonical binding features. After functional enrichment, seven new targets were genes associated with endocrine resistance (*p*-value < 1.28e-02) and five genes were associated to endocrine and calcium reabsorption (*p*-value < 1.64e-02) categories. At least three genes involved in new microRNA\*:target interactions (CDKN1B, NOTCH2, CCND1) are consistent with current gene-disease association literature related to their downregulation in prostate cancer progression. These results support the central hypothesis of this thesis, as microRNA gene variability interferes with the complementarity of microRNA:mRNA, perturbing target recognition leading to the downregulation of genes associated to disease susceptibility.

One of the biggest challenges for computational studies is to predict a phenotype as accurately as possible and to reconcile results with high-throughput experimental approaches such as microarrays, RNA-seq or AGO-PAR-CLIP in case of microRNA:target interactions. In order to provide evidence of gene expression changes due to mutated microRNAs, a microarray expression analysis of PC3 versus prostate epithelial cells was performed using untreated samples using the hgu133plus2 platform from the GEO database. A low number of expected targets were observed to be underexpressed in PC3, probably due to the small sample size of data sets available at GEO. Hence, a Fisher's exact test was used to translate the frequency of perturbed targets into a *p*-value for the assessment of statistically significance. Only microRNA-4804\*-5p observed underexpressed targets TNKS and GUCY1A3 were statistically significant (*p*-value < 0.05). However, neither TNKS nor GUCY1A3 were in the group of genes previously found to be enriched for endocrine resistance via WebGestalt. This indirect method assessed the effect

of microRNAs<sup>\*</sup> over predicted target expression assuming the untreated PC3 samples have enough microRNA expression abundance for repression to be detectable<sup>91</sup>. Due to the lack of pairwise microRNA and gene expression data in PC3, we cannot completely predict the effect of our candidate microRNAs<sup>\*</sup> solely with this indirect experimental approach.

Furthermore, four groups of genes assessed by the hgu133plus2 platform were organized based on their predicted interactions with non-mutated microRNA and mutated microRNA. A Kruskal-Wallis non-parametric test was performed in order to assess for statistical significance between groups, where Group A (genes not subject to the regulation of the present study microRNA candidates) versus Group B (genes regulated by candidate non-mutated microRNAs), and Group A versus Group C (genes regulated by candidate mutated microRNAs) showed differences regarding group medians. It would have been interesting to observe statistical significance comparing Group B versus Group C, but with AGO-PAR-CLIP experimental data to assess the effect of SNPs affecting the seed region of miR-3620\*-5p, miR-3161\*-5p, miR-1178\*-5p, miR-4804\*-5p, and miR-449c\*-3p.

A third statistical assessment was performed by querying twelve candidate microRNA\*:target interactions summarized in Table 4 against the CancerMiner database. CancerMiner is an online tool from The Cancer Genome Atlas (TCGA) which evaluates recurrence of miRNA-mRNA expression association across ten epithelial cancer types from TCGA<sup>135</sup>. This method considers pairwise miRNA-mRNA relationships in individual cancer types and evaluates these relationships using a multivariate linear model. Associations are rank transformed in individual cancer types, and subsequently tested with the null hypothesis that no association exists between the miRNA-mRNA pair in all cancer types<sup>14</sup>. Only the relationship of miRNA-3161:CDKN1B was retrieved as a recurrent expression association in uterine corpus endometrial carcinoma (UCEC) with an association score (REC) by individual cancer type of -0.44 and a *p*-value of 0.366. Although the REC score did not provide a very strong association at the individual level, the result supports that miRNA-3161:CDKN1B interaction is likely to be

functionally relevant in different cancer types. In the context of this study, results suggest that CDKN1B dysregulation by miR-3161\*-5p would be leading to PC3 proliferation due to the lack of cell cycle arrest from phase G1 to S (Figure 18). CDKN1B (p27<sup>Kip1</sup>) is a member of the cip/kip family of cyclin-dependent kinase inhibitors that phosphorylate the retinoblastoma protein (Rb) through binding to cyclin-cdk complexes: cyclin E-cdk2; cyclin A-cdk2; and cyclin D-cdk4<sup>173,174</sup>. Failure of cell cycle arrest secondary to alterations in cdk inhibitor expression has been implicated in prostate cancer progression<sup>175,176</sup> and luminal breast cancer (LBC)<sup>177</sup>. CDKN1B has been found to be mutated in few other neoplasia such as small intestine neuroendocrine tumors, a rare carcinoma arising from endocrine precursor cells in which CDKN1B represents the most frequently mutated gene<sup>139</sup>. In addition, CDKN1B germline mutations have been also proposed to be the cause of Multiple Endocrine Neoplasia type 4 (MEN4), an autosomal dominant disorder characterized by the occurrence of tumors in endocrine glands<sup>178</sup>.

Prostate cancer cell line PC3 has been described in thousands of publications over the last four decades and has shown to share features with prostatic small cell neuroendocrine carcinomas (SCNC) with the implication that molecular mechanisms and therapeutic efficacies observed with PC3 cells are likely applicable to SCNC<sup>1</sup>. Prostatic small cell neuroendocrine carcinoma is a variant form of prostate cancer often characterized by an aggressive course with a poor response to conventional androgen deprivation therapy (ADT), consistent with the lack of the androgen receptor in prostatic small cell carcinoma (SCC)<sup>2</sup>. In some men treated with ADT, development of small cell carcinoma might represent the "escape" of a subpopulation of hormone-independent cells resulting from the selective pressure of hormonal therapy<sup>3</sup>. This phenomenon has been observed in other hormone-driven cancer types such as human basal like breast cancer cells with the downregulation of p27<sup>Kip1</sup> which abrogates antiestrogen-mediated cell cycle arrest<sup>179</sup>. Hence, the suggestion of CDKN1B dysregulation by miR-3161\*-5p might go beyond the idiosyncrasy of the PC3 cell line, but rather an interesting future direction to investigate prostate

cancer patients with small cell neuroendocrine carcinoma rendering to an adverse disease outcome due to uncontrolled cell proliferation.



**Figure 18** CDKN1B dysregulation by miR-3161\*-5p might lead to PC3 proliferation due to the lack of cell cycle arrest from phase G1 to S. Molecular mechanisms and therapeutic efficacies observed with PC3 cells are likely applicable to small cell neuroendocrine carcinoma (SCNC), a variant form of prostate cancer often characterized by an aggressive course with a poor response to conventional androgen deprivation therapy (ADT). The pathway from CDKN1B to E2DFP was obtained from KEGG.

Copyright releases of all previously published figures that have been used in this thesis, either unmodified or without substantial modification, have been requested/obtained.

These are kept in: https://github.com/microRNA-PCa/PC3

# References

- 1. Tai, S. *et al.* PC3 Is a Cell Line Characteristic of Prostatic Small Cell Carcinoma. *Prostate* **71**, 1668–1679 (2011).
- Grigore, A. D., Ben-Jacob, E. & Farach-Carson, M. C. Prostate Cancer and Neuroendocrine Differentiation: More Neuronal, Less Endocrine? *Front. Oncol.* (2015). doi:10.3389/fonc.2015.00037
- 3. Spiess, P. E. *et al.* Treatment outcomes of small cell carcinoma of the prostate: A singlecenter study. *Cancer* **110**, 1729–1737 (2007).
- 4. Prostate Cancer Statistical Facts. *National Cancer Institute The Surveillance Epidemiology and End Results (SEER) Program* (2014). Available at: https://seer.cancer.gov/statfacts/html/prost.html. (Accessed: 11th March 2018)
- 5. Mayo Clinic. Prostate cancer: Diagnosis and treatment. (2018). Available at: https://www.mayoclinic.org/diseases-conditions/prostate-cancer/diagnosis-treatment/drc-20353093. (Accessed: 11th March 2018)
- Sartor, O., Eisenberger, M., Kattan, M. W., Tombal, B. & Lecouvet, F. Unmet Needs in the Prediction and Detection of Metastases in Prostate Cancer. *Oncologist* (2013). doi:10.1634/theoncologist.2013-0027
- 7. Jakobsen, N. A., Hamdy, F. C. & Bryant, R. J. Novel biomarkers for the detection of prostate cancer. *J. Clin. Urol.* **9**, 3–10 (2016).
- 8. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).
- 9. Hayes, J., Peruzzi, P. P. & Lawler, S. *MicroRNAs in cancer: biomarkers, functions and therapy. Trends in Molecular Medicine* **20**, 460–469 (2014).
- 10. Razvi, E. miRNA Therapeutics Heading to the Clinic. *Genet. Eng. Biotechnol. News* **34**, 12–12 (2014).
- 11. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866 (2006).
- 12. Kasinski, A. L. & Slack, F. J. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer* **11**, 849–864 (2011).
- 13. Stahlhut, C. & Slack, F. J. MicroRNAs and the cancer phenotype: profiling, signatures and clinical implications. *Genome Med.* **5**, 111 (2013).
- 14. Jacobsen, A. *et al.* Analysis of microRNA-target interactions across diverse cancer types. **20**, 1325–1332 (2013).
- 15. Yang, D. *et al.* Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer Cell* **23**, 186–199 (2014).
- 16. Wu, H. Q. *et al.* Flux balance analysis predicts Warburg-like effects of mouse hepatocyte deficient in miR-122a. *PLoS Comput. Biol.* **13**, 1–22 (2017).
- 17. Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of

mammalian microRNA targets. Cell 115, 787–98 (2003).

- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* 120, 15–20 (2005).
- 19. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cell* **27**, 91–105 (2007).
- 20. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–11 (2004).
- 21. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* **64**, 320–333 (2016).
- 22. Bracken, C. P., Scott, H. S. & Goodall, G. J. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat. Publ. Gr.* **17**, (2016).
- 23. Stegeman, S. *et al.* A Large-Scale Analysis of Genetic Variants within Putative miRNA Binding Sites in Prostate Cancer. *Cancer Discov.* **5**, 368–379 (2015).
- 24. Meister, G. *et al.* Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
- 25. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–41 (2004).
- Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the Seed-Target Recognition Step in RNA Silencing by a PIWI/MID Domain Protein. *Mol. Cell* 33, 204–214 (2009).
- Wee, L. M. M., Flores-Jasso, C. F. F., Salomon, W. E. E. & Zamore, P. D. D. Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055–1067 (2012).
- 28. Ghanbari, M. *et al.* Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. (2016). doi:10.1038/srep28387
- 29. Ryan, B. M., Robles, A. I. & Harris, C. C. *Genetic variation in microRNA networks: The implications for cancer research. Nature Reviews Cancer* **10**, 389–402 (2010).
- 30. Seim, I., Jeffery, P. L., Thomas, P. B., Nelson, C. C. & Chopin, L. K. Whole-Genome Sequence of the Metastatic PC3 and LNCaP Human Prostate Cancer Cell Lines. *G3; Genes/Genomes/Genetics* **7**, g3.117.039909 (2017).
- 31. Kirby, R. S. & Patel, M. I. *Fast Facts: Prostate Cancer*. (Oxford: Health Press, 2017). doi:9781461952633. 9781908541543.
- 32. Brawley, O. W. Prostate cancer epidemiology in the United States. *World Journal of Urology* **30**, 195–200 (2012).
- City of Hope Staff. Prostate Cancer Symptoms, Diagnosis, Treatment & amp; Research | City of Hope, Los Angeles, CA. Available at: https://www.cityofhope.org/clinicalprogram/prostate-cancer. (Accessed: 24th March 2018)
- 34. Haiman, C. A. *et al.* Characterizing genetic risk at known prostate cancer susceptibility loci in African Americans. *PLoS Genet.* **7**, e1001387 (2011).

- 35. Shikany, J. M. *et al.* Dietary carbohydrate, glycemic index, glycemic load, and risk of prostate cancer in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) cohort. *Cancer Causes Control* **22**, 995–1002 (2011).
- 36. Masko, E. M. *et al.* Evidence for Feedback Regulation Following Cholesterol Lowering Therapy in a Prostate Cancer Xenograft Model. *Prostate* **77**, 446–457 (2017).
- 37. Sandhu, J. S. *Urinary Dysfunction in Prostate Cancer*. (Springer International Publishing, 2016). doi:10.1007/978-3-319-23817-3
- Giunchi, F. *et al.* Interpathologist concordance in the histological diagnosis of focal prostatic atrophy lesions, acute and chronic prostatitis, PIN, and prostate cancer. *Virchows Arch.* 470, 711–715 (2017).
- 39. Al-mamari, S. A. & Al-busaidy, S. S. *Urological Cancer Management*. (Springer International Publishing, 2015). doi:10.1007/978-3-319-16301-7
- 40. Walsh, A. L., Considine, S. W., Thomas, A. Z., Lynch, T. H. & Manecksha, R. P. Digital rectal examination in primary care is important for early detection of prostate cancer: a retrospective cohort analysis study. *Br. J. Gen. Pract.* **64**, e783–e787 (2014).
- 41. Grall, J. & Corbel, L. PSA et hypertrophie bénigne de la prostate. *Ann. Urol. (Paris).* **38,** S43–S45 (2004).
- 42. McVary, K. T. *et al.* Management of Benign Prostatic Hyperplasia (BPH) Guideline. *AUA Clin. Guidel.* (2016).
- 43. Crawford, E. D. & Abrahamsson, P. A. PSA-based Screening for Prostate Cancer: How Does It Compare with Other Cancer Screening Tests? *Eur. Urol.* **54**, 262–273 (2008).
- 44. Schröder, F. H. Screening for prostate cancer (PC)-an update on recent findings of the European Randomized Study of Screening for Prostate Cancer (ERSPC). *Urol. Oncol. Semin. Orig. Investig.* **26**, 533–541 (2008).
- 45. Lin, D. W. Beyond PSA: Utility of novel tumor markers in the setting of elevated PSA. *Urologic Oncology: Seminars and Original Investigations* **27**, 315–321 (2009).
- 46. Carter, H. B. *et al.* Early Detection of Prostate Cancer Guideline. *AUA Clin. Guidel.* 1–28 (2013).
- Sanda, M. *et al.* Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. *AUA Clinical Guidelines* (2017). Available at: http://www.auanet.org/guidelines/clinically-localized-prostate-cancer-new-(aua/astro/suo-guideline-2017). (Accessed: 20th January 2018)
- 48. Bostwick, D. G. & Meiers, I. Neoplasms of the prostate. in *Urologic Surgical Pathology Third Edition* Chapter 9: 408-531 (Elsevier Health Sciences, 2014).
- 49. Sedelaar, J. P. M. & Schalken, J. A. The need for a personalized approach for prostate cancer management. *BMC medicine* **13**, 109 (2015).
- 50. Ablin, R. J. & Mason, M. D. *Metastasis of prostate cancer. Springer* (2008). doi:10.1007/s13398-014-0173-7.2
- 51. Zhou, Y., Bolton, E. C. & Jones, J. O. Androgens and androgen receptor signaling in prostate tumorigenesis. *J. Mol. Endocrinol.* **54**, R15–R29 (2015).

- 52. Uemura, M. *et al.* Novel 5α-steroid reductase (SRD5A3, type-3) is overexpressed in hormone-refractory prostate cancer. *Cancer Sci.* **99**, 81–86 (2008).
- 53. Titus, M. A. *et al.* 5alpha-ReductaseType 3 Enzyme in Benign and Malignant Prostate. *Prostate* **74**, 235–249 (2014).
- 54. Thomas, L. N. *et al.* Differential alterations in 5alpha-reductase type 1 and type 2 levels during development and progression of prostate cancer. *Prostate* **63**, 231–239 (2005).
- 55. Amory, J. K. *et al.* The effect of 5alpha-reductase inhibition with dutasteride and finasteride on bone mineral density, serum lipoproteins, hemoglobin, prostate specific antigen and sexual function in healthy young men. *J. Urol.* **179**, 2333–2338 (2008).
- 56. Harris, W. P., Mostaghel, E. A., Nelson, P. S. & Montgomery, B. Androgen deprivation therapy: Progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nature Clinical Practice Urology* **6**, 76–85 (2009).
- 57. Nadal, R., Schweizer, M., Kryvenko, O. N., Epstein, J. I. & Eisenberger, M. A. Small cell carcinoma of the prostate. *Nat Rev Urol* **1121**, 213–219 (2014).
- 58. Beltran, H. *et al.* Aggressive variants of castration-resistant prostate cancer. *Clinical Cancer Research* **20**, 2846–2850 (2014).
- 59. Kar, S. P. *et al.* Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discov.* **6**, 1052–1067 (2016).
- 60. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
- 61. Freedman, M. L. *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics* **43**, 513–518 (2011).
- 62. Goh, C. L. *et al.* Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *Journal of Internal Medicine* **271**, 353–365 (2012).
- 63. Willenbrock, H. *et al.* Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *RNA* **15**, 2028–34 (2009).
- 64. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- 65. Zhao, S. *et al.* QuickMIRSeq: A pipeline for quick and accurate quantification of both known miRNAs and isomiRs by jointly processing multiple samples from microRNA sequencing. *BMC Bioinformatics* **18**, (2017).
- 66. Kote-Jarai, Z. *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat. Publ. Gr.* **43**, (2011).
- 67. Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025 (2015).
- 68. Thomas, P. D. & Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. *Proc. Natl. Acad. Sci.* **101**, 15398–15403 (2004).
- 69. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human

transcriptome. Nature 505, 706–709 (2014).

- 70. Ghanbari, M. *et al.* Genetic variants in microRNAs and their binding sites within gene 3'UTRs associate with susceptibility to age-related macular degeneration. *Hum. Mutat.* **38**, 827–838 (2017).
- 71. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* **11**, 597–610 (2010).
- 72. Brownawell, A. M. & Macara, I. G. Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J. Cell Biol.* **156**, 53–64 (2002).
- 73. Melo, S. A. *et al.* A genetic defect in exportin-5 traps precursor MicroRNAs in the nucleus of cancer cells. *Cancer Cell* **18**, 303–315 (2010).
- 74. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* **15**, 321–333 (2015).
- 75. Baek, D. *et al.* The impact of microRNAs on protein output. (2008). doi:10.1038/nature07242
- 76. Catalanotto, C., Cogoni, C. & Zardo, G. MicroRNA in control of gene expression: An overview of nuclear functions. *Int. J. Mol. Sci.* **17**, (2016).
- 77. Zander, A., Holzmeister, P., Klose, D., Tinnefeld, P. & Grohmann, D. Single-molecule FRET supports the two-state model of argonaute action. *RNA Biol.* **11**, 45–56 (2014).
- 78. Iwasaki, S. *et al.* Defining fundamental steps in the assembly of the Drosophila RNAi enzyme complex. *Nature* **521**, 533–536 (2015).
- Salomon, W. E. E., Jolly, S. M. M., Moore, M. J. J., Zamore, P. D. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* 162, 84–95 (2015).
- Yao, C., Sasaki, H. M., Ueda, T., Tomari, Y. & Tadakuma, H. Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex. *Mol. Cell* 59, 125–132 (2015).
- 81. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599–606 (2004).
- 82. Rivas, F. V *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **12**, 340–349 (2005).
- 83. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. *Cell* **130**, 101–112 (2007).
- 84. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–7 (2004).
- 85. Elkayam, E. *et al.* The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110 (2012).
- 86. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. *Nature* **486**, 368–374 (2012).
- 87. Schirle, N. T. & MacRae, I. J. The crystal structure of human argonaute2. Science (80-.).

**336**, 1037–1040 (2012).

- 88. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science (80-. ).* **346,** 608–613 (2014).
- 89. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nature Structural and Molecular Biology* **21**, 743–753 (2014).
- 90. Miyoshi, T., Ito, K., Murakami, R. & Uchiumi, T. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nat. Commun.* **7**, (2016).
- 91. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA targets in mammalian mRNAs.pdf. *Elife* (2015). doi:10.7554/eLife.05005.001
- 92. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol.* 5, R1 (2003).
- 93. Dweep, H., Sticht, C., Pandey, P. & Gretz, N. MiRWalk Database: Prediction of possible miRNA binding sites by " walking" the genes of three genomes. *J. Biomed. Inform.* **44**, 839–847 (2011).
- 94. Sethupathy, P., Megraw, M. & Hatzigeorgiou, A. G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods* **3**, 881–886 (2006).
- 95. Choudhury, A. D. *et al.* The role of genetic markers in the management of prostate cancer. *European Urology* **62**, 577–587 (2012).
- 96. Dweep, H. & Gretz, N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nature methods* **12**, 697 (2015).
- 97. Hamilton, M. P. *et al.* The Landscape of microRNA Targeting in Prostate Cancer Defined by AGO-PAR-CLIP. *Neoplasia (United States)* **18**, 356–370 (2016).
- 98. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. (2005).
- 99. Pichler, M. & Calin, G. A. MicroRNAs in cancer: from developmental genes in worms to their clinical application in patients. *Br. J. Cancer* **113**, 569–573 (2015).
- 100. Di Leva, G., Garofalo, M. & Croce, C. M. MicroRNAs in cancer. *Annu. Rev. Pathol.* **9,** 287–314 (2014).
- 101. Kumar, B. & Lupold, S. E. MicroRNA expression and function in prostate cancer: a review of current knowledge and opportunities for discovery. *Asian J. Androl.* **18**, 559–67 (2016).
- 102. Lo, U.-G., Yang, D. & Hsieh, J.-T. The role of microRNAs in prostate cancer progression. *Transl. Androl. Urol.* **2**, 228–41 (2013).
- 103. Bao, B.-Y. *et al.* Polymorphisms inside microRNAs and microRNA target sites predict clinical outcomes in prostate cancer patients receiving androgen-deprivation therapy. *Clin. Cancer Res.* **17**, 928–36 (2011).
- 104. Lin, R.-J. *et al.* microRNA signature and expression of Dicer and Drosha can predict prognosis and delineate risk groups in neuroblastoma. *Cancer Res.* **70**, 7841–50 (2010).
- 105. Merritt, W. M. *et al.* Dicer, Drosha, and Outcomes in Patients with Ovarian Cancer. *n engl j med* **35925**, (2008).

- 106. Guo, X. *et al.* The microRNA-processing enzymes: Drosha and Dicer can predict prognosis of nasopharyngeal carcinoma. *J. Cancer Res. Clin. Oncol.* **138**, 49–56 (2012).
- 107. Ashley, V. A. *et al.* The Use of Biomarkers in Prostate Cancer Screening and Treatment. *Rev. Urol.* **19**, 221–234 (2017).
- 108. Kristiansen, G. Markers of clinical utility in the differential diagnosis and prognosis of prostate cancer. *Mod. Pathol.* **31**, S143-155 (2018).
- Chin, L. J. *et al.* A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res.* 68, 8535–8540 (2008).
- 110. Jazdzewski, K. *et al.* Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc. Natl. Acad. Sci.* **105**, 7269–7274 (2008).
- 111. Mencía, A. *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat. Genet.* **41**, 609–613 (2009).
- 112. NCBI Handout Series. dbSNP : Database for Short Genetic Variations. *Fact Sheet* 2 (2014).
- 113. Roche. Hybridization Capture-Based Target Enrichment for NGS Roche Sequencing Solutions. Available at: http://sequencing.roche.com/en/products-solutions/by-category/target-enrichment/hybridization.html. (Accessed: 4th April 2018)
- 114. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
- 115. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
- 116. La Jolla California USA. GraphPad Software. Available at: https://www.graphpad.com/.
- 117. ATCC. Prostate cancer cell line PC3 (CRL-1435). Available at: https://www.atcc.org/en/Products/All/CRL-1435.aspx#characteristics.
- 118. ATCC, P. S. Pc-3 (atcc ® crl-1435 <sup>™</sup>). 1–3
- 119. ScienCell Research Laboratories. Human Prostate Epithelial Cell genomic DNA. Available at: https://www.sciencellonline.com/human-prostate-epithelial-cell-genomic-dna.html.
- 120. ScienCell Research Laboratories. HPrEpiC gDNA ScienCell Catalog Number 4419.
- 121. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: A fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
- 122. SoftGenetics, L. NextGENe User's Manual. 464 (2016).
- 123. Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, (2015).
- 124. Griffiths-Jones, S. miRBase: the microRNA sequence database. *Methods Mol. Biol.* **342**, 129–38 (2006).

- 125. Bartel, D. P. TargetScan 7 UTR profiles. *Whitehead Institute for Biomedical Research* 1 (2016). Available at: http://www.targetscan.org/docs/UTR\_profiles.html.
- 126. Nam, J.-W. W. *et al.* Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol. Cell* **53**, 1031–1043 (2014).
- 127. Feng, X., Li, L., Wagner, E. J. & Li, W. TC3A: The Cancer 3 UTR Atlas. *Nucleic Acids Res.* **46**, (2018).
- 128. Mayr, C. & Bartel, D. P. Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**, 673–684 (2009).
- 129. Ghent University. BEG Venn Diagram Tool.
- 130. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741–W748 (2005).
- 131. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
- 132. Hu, J. X., Zhao, H. & Zhou, H. H. False discovery rate control with groups. *J. Am. Stat. Assoc.* **105**, 1215–1227 (2010).
- 133. Irizarry, R. ormalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
- 134. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- 135. Jacobsen, A. *et al.* Supplemental Information -Analysis of microRNA-target interactions across diverse cancer types. *Nat. Struct. Mol. Biol.* **20**, 1325–1332 (2013).
- 136. Chi, S. W., Hannon, G. J. & Darnell, R. B. An alternative mode of microRNA target recognition. *Nat. Struct. Mol. Biol.* **19**, 321–327 (2012).
- 137. Noe, J. *et al.* Impaired expression and function of the bile salt export pump due to three novel ABCB11 mutations in intrahepatic cholestasis. *J. Hepatol.* **43**, 536–543 (2005).
- 138. Ho, R. H. *et al.* Polymorphic variants in the human bile salt export pump (BSEP; ABCB11): functional characterization and interindividual variability. *Pharmacogenet. Genomics* **20**, 45–57 (2010).
- 139. Francis, J. M. *et al.* Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nat. Genet.* **45**, 1483–1486 (2013).
- 140. Kibel, A. S. *et al.* CDKN1A and CDKN1B polymorphisms and risk of advanced prostate carcinoma. *Cancer Res* **63**, 2033–2036 (2003).
- 141. Molatore, S. *et al.* A novel germline CDKN1B mutation causing multiple endocrine tumors: clinical, genetic and functional characterization. *Hum Mutat* **31**, E1825-35 (2010).
- 142. Canbay, E. *et al.* CCND1 and CDKN1B polymorphisms and risk of breast cancer. *Anticancer Res.* **30**, 3093–3098 (2010).
- 143. Pinchot, S. N. *et al.* Identification and validation of Notch pathway activating compounds through a novel high-throughput screening method. *Cancer* **117**, 1386–1398 (2011).

- 144. Garcia, A. & Kandel, J. J. Notch: A key regulator of tumor angiogenesis and metastasis. *Histology and Histopathology* **27**, 151–156 (2012).
- 145. Su, Q. & Xin, L. Notch signaling in prostate cancer: Refining a therapeutic opportunity. *Histology and Histopathology* **31**, 149–157 (2016).
- 146. Leung, K. K. *et al.* Enhanced Prediction of Src Homology 2 (SH2) Domain Binding Potentials Using a Fluorescence Polarization-derived c-Met, c-Kit, ErbB, and Androgen Receptor Interactome. *Mol. Cell. Proteomics* **13**, 1705–1723 (2014).
- 147. Jones, N. *et al.* Analysis of a Shc Family Adaptor Protein, ShcD/Shc4, That Associates with Muscle-Specific Kinase. *Mol. Cell. Biol.* **27**, 4759–4773 (2007).
- 148. Costa, C. *et al.* A novel ATP1A2 gene mutation in familial hemiplegic migraine and epilepsy. *Cephalalgia* **34**, 68–72 (2014).
- 149. Gallanti, A. *et al.* A novel de novo nonsense mutation in ATP1A2 associated with sporadic hemiplegic migraine and epileptic seizures. *J. Neurol. Sci.* **273**, 123–126 (2008).
- 150. Thomsen, L. L. *et al.* Screen for CACNA1A and ATP1A2 mutations in sporadic hemiplegic migraine patients. *Cephalalgia* **28**, 914–921 (2008).
- 151. Wang, Y.-M. *et al.* ESR1 Gene Polymorphisms and Prostate Cancer Risk: A HuGE Review and Meta-Analysis. *PLoS One* **8**, e66999 (2013).
- 152. Holst, F. *et al.* Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nat. Genet.* **39**, 655–660 (2007).
- 153. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45**, 1446–1451 (2013).
- 154. Tsai, H. *et al.* Cyclin D1 Loss Distinguishes Prostatic Small-Cell Carcinoma from Most Prostatic Adenocarcinomas. *Clin. Cancer Res.* **21**, 5619–29 (2015).
- 155. Nakamura, Y. *et al.* Cyclin D1 (CCND1) expression is involved in estrogen receptor beta (ERβ) in human prostate cancer. *Prostate* **73**, 590–595 (2013).
- 156. Mavaddat, N., Dunning, A. M., Ponder, B. A. J., Easton, D. F. & Pharoah, P. D. Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **18**, 255–259 (2009).
- 157. Watanabe, G. *et al.* Inhibition of cyclin D1 kinase activity is associated with E2F-mediated inhibition of cyclin D1 promoter activity through E2F and Sp1. *Mol. Cell. Biol.* **18**, 3212–22 (1998).
- 158. Sankpal, U. T., Goodison, S., Abdelrahim, M. & Basha, R. Targeting SP1 Transcription Factor in Prostate Cancer Therapy. *Med. Chem. (Los. Angeles).* **7**, 518–525 (2011).
- 159. Choi, J. A., Jung, Y. S., Kim, J. Y., Kim, H. M. & Lim, I. K. Inhibition of breast cancer invasion by TIS21/BTG2/Pc3-Akt1-Sp1-Nox4 pathway targeting actin nucleators, mDia genes. *Oncogene* **35**, 83–93 (2016).
- 160. Bajpai, R. & Nagaraju, G. P. Specificity protein 1: Its role in colorectal cancer progression and metastasis. *Critical Reviews in Oncology/Hematology* **113**, 1–7 (2017).
- 161. Côté, M., Guillon, G., Payet, M. D. & Gallo-Payet, N. Expression and regulation of adenylyl cyclase isoforms in the human adrenal gland. *J. Clin. Endocrinol. Metab.* **86**, 4495–503
(2001).

- 162. Bogard, A. S., Xu, C. & Ostrom, R. S. Human bronchial smooth muscle cells express adenylyl cyclase isoforms 2, 4, and 6 in distinct membrane microdomains. *J. Pharmacol. Exp. Ther.* **337**, 209–17 (2011).
- 163. Teh, C. E. *et al.* Heterozygous mis-sense mutations in Prkcb as a critical determinant of anti-polysaccharide antibody formation. *Genes Immun.* **14**, 223–233 (2013).
- 164. Cook, P. J. *et al.* Somatic chromosomal engineering identifies BCAN-NTRK1 as a potent glioma driver and therapeutic target. *Nat. Commun.* **8**, (2017).
- 165. Kelly, B. T. *et al.* AP2 controls clathrin polymerization with a membrane-activated switch. *Science (80-. ).* **345**, 459–463 (2014).
- 166. Rappoport, J. Z., Taha, B. W., Lemeer, S., Benmerah, A. & Simon, S. M. The AP-2 Complex Is Excluded from the Dynamic Population of Plasma Membrane-associated Clathrin. *J. Biol. Chem.* **278**, 47357–47360 (2003).
- 167. Grass, B. *et al.* Immunohistochemical detection of EGFR, fibrillin-2, P-cadherin and AP2β as biomarkers for rhabdomyosarcoma diagnostics. *Histopathology* **54**, 873–879 (2009).
- 168. Kepp, K. *et al.* Hypervariable intronic region in NCX1 is enriched in short insertion-deletion polymorphisms and showed association with cardiovascular traits. *BMC Med. Genet.* **11**, (2010).
- 169. Arking, D. E. *et al.* Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* **46**, 826–836 (2014).
- GeneCards. TNKS. Available at: https://www.genecards.org/Search/Keyword?queryString=TNKS.
- 171. GeneCards. GUCY1A3 related genes. Available at: https://www.genecards.org/Search/Keyword?queryString=GUCY1A3.
- 172. Buckley, A. R. *et al.* Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC Genomics* **18**, (2017).
- 173. Polyak, K. *et al.* Cloning of p27Kip1, a Cyclin-Dependent Kinase Inhibitor and a Potential Mediator of Extracellular Antimitogenic Signals. *Cell* **76**, 59–66 (1994).
- 174. Elledge, S. J. Cell cycle checkpoints: Preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
- 175. Cote, R. J., Shi, Y., Feng, A.-C., Skinner, D. & Lieskovosky, G. Association of p27 Kip1 Levels With Recurrence and Survival in Patients With Stage C Prostate Carcinoma Patients, Materials, and. **90**, (1998).
- 176. Cordon-Cardo, C. *et al.* Distinct altered patterns of p27(KIP1) gene expression in benign prostatic hyperplasia and prostatic carcinoma. *J. Natl. Cancer Inst.* **90**, 1284–1291 (1998).
- 177. Ignatiadis, M. & Sotiriou, C. Luminal breast cancer: From biology to treatment. *Nature Reviews Clinical Oncology* **10**, 494–506 (2013).
- 178. Lee, M. & Pellegata, N. S. Multiple endocrine neoplasia type 4. *Frontiers of Hormone Research* **41**, 63–78 (2013).

179. Cariou, S. *et al.* Down-regulation of p21WAF1/CIP1 or p27Kip1 abrogates antiestrogenmediated cell cycle arrest in human breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9042–9046 (2000).

## **Appendix**

## Supplementary Table 1:

From 546 predicted genes interacting with non-mutated microRNA, only 510 had associated Entrez gene IDs which were considered by WebGestalt for the ORA analysis. Table 4 summarizes the ORA results of genes from gained microRNA:target interactions.

ID	Name	#Gene
<u>hsa05218</u>	Melanoma - Homo sapiens (human)	6
<u>hsa01522</u>	Endocrine resistance - Homo sapiens (human)	7
<u>hsa04070</u>	Phosphatidylinositol signaling system - Homo sapiens (human)	7
hsa05223	Non-small cell lung cancer - Homo sapiens (human)	5
<u>hsa04144</u>	Endocytosis - Homo sapiens (human)	13
<u>hsa04020</u>	Calcium signaling pathway - Homo sapiens (human)	10
<u>hsa05033</u>	Nicotine addiction - Homo sapiens (human)	4
<u>hsa05219</u>	Bladder cancer - Homo sapiens (human)	4
<u>hsa00534</u>	Glycosaminoglycan biosynthesis - heparan sulfate / heparin - Homo sapiens (human)	3
hsa05215	Prostate cancer - Homo sapiens (human)	6

Summary of the enriched categories of 510 Entrez gene IDs out 546 predicted genes interacting with miR-3620-5p, miR-3161-5p, miR-1178-5p, miR-4804-5p, and miR-449c-3p without a SNP mapping within the seed sequence region. The first column ID corresponds to the KEGG pathway ID annotated in the human species (hsa).