



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2014

Identification of micro-RNAs and their messenger RNA targets in Prostate cancer and Biological fluids

Kanika Sharma
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Bioinformatics Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/3551>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Kanika Sharma

2014 All Rights Reserved

Identification of micro-RNAs and their messenger RNA targets in Prostate cancer and Biological fluids

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science
in Bioinformatics at Virginia Commonwealth University.

by

Kanika Sharma

B.Sc. Bioinformatics

Virginia Commonwealth University May, 2012

Director: Zendra Zehner, PhD,

Professor, Department of Biochemistry and Molecular Biology

Virginia Commonwealth University

Richmond, Virginia

August, 2014

DEDICATED TO MY PARENTS
Mr. Pramod Kumar Sharma and Mrs. Vimala Sharma

Acknowledgement

I feel extremely blessed to be surrounded by some of the most influential people around me. I would like to first extend my gratitude to my parents, who have stuck with me through thick and thin and have encouraged me to stay strong and positive throughout my graduate journey. I sincerely thank you for everything ma and pa.

My bioinformatics journey began when I was a sophomore in college, who knew back then that one day I will receive my M.Sc. in this field. I still remember the first day I met Dr. Allison Johnson (the God of Bioinformatics). I went into her office, without any appointment, to ask what exactly was BNFO 301 and why was it so confusing. Ever since then she has been an important part of my academic life and has provided me with ample opportunities to make bioinformatics a fun experience of my life. I will never forget the guidance you have showered upon me and all the help you provided me to finish my thesis successfully.

I am deeply indebted to my dearest advisor Dr. Zendra Zehner, for taking a chance on me and giving me the opportunity to work in her lab as well as work on this project. I honestly, cannot thank you enough for the immense knowledge and support you have provided me, both academically and personally. I can never forget the times when I used to get disappointed with my negative results, and you always found a way to make it look positive and less disappointing. You have always been there when I needed your help the most, not matter how busy you were, you always found time to help me. I know I must have troubled you a lot over this one year, yet, despite my many shortcomings, you were always patient with me and encouraged me to do my best. This shows what a wonderful person you are and that you really care for your students. You are truly the best advisor I have gotten so far. I sincerely thank you from the bottom of my heart for making this one year so memorable for me.

I would also like to thank Dr. Sarah Seashols, who took time out, while she was pursuing her Ph.D., to teach me all the lab work. She was always ready to help and bring out the best in me. She is an excellent teacher and even more a very soft hearted person. I remember during the computational analyses when things were not working out as we had anticipated, she stuck with me and encouraged me to keep looking for answers. She well understood the task was not easy, yet she showed immense faith in me. I think it is all because of her that I was able to finish the computational analyses on time and come out with some great results. I thank you too, from the bottom of my heart for constantly supporting me and editing my work, despite being busy with other commitments.

I will never forget my jovial committee member, Dr. Mike Holmes who was always there to lighten up the surrounding, by finding funny jokes to crack, when things got serious and worrisome. He was always present to provide encouragement when I need it the most. He also showed concern about my future plans and helped me find the best routes to pursue my goals. I thank you from the bottom of my heart for your invaluable advice and suggestions on how to become successful in life.

I would like to thank Dr. William Budd who introduced me to the Zehner lab and its wonderful people. It was because of him and Dr. Zehner, that I got the opportunity work on this important project. He was constantly present to teach me the lab work and help me plan out the research for my masters. I thank you again, for helping me get started.

Finally, I thank my department advisor Dr. Herschel Emery, my lab-mates, all my professors, and all my friends who have motivated me and helped me get to where I am today.

Table of Contents

	Page
List of figures.....	viii
List of tables.....	x
Abbreviations.....	xi
Abstract.....	xiii
Chapter 1- Introduction, background and foundation of miRNA research.....	1
1.1 Prostate cancer and its origin.....	2
1.2 Common of prostate cancer.....	3
1.3 Screening of prostate cancer.....	4
1.4 Treatments of prostate cancer.....	7
1.5 Identifying biomarkers for prostate cancer.....	9
1.6 Micro-RNAs and their role in oncogenesis.....	9
1.7 Biogenesis of miRNAs.....	10
1.8 Nomenclature of miRNAs.....	13
1.9 Research objectives, previous work, and models used.....	13
1.9.1 Prostate cancer progression model M12 cell line.....	14
1.9.2 Other prostate cancer cell lines.....	16
1.9.3 Identification of dysregulated miRNAs through P69-M12 model cell lines.....	16
1.10 Selection of miRNAs through screening processes.....	18
1.11 The purpose of this thesis.....	26
Chapter 2- Elucidation and confirmation of miR-17-3p binding sites within 3'UTR of ErbB2.....	29
2.1 History of miR-17-3p and its target vimentin.....	30
2.2 Finding additional 3'UTR and mRNA targets for miR-17-3p.....	31
2.3 MiR-17-3p a potential target for 3'UTR ErbB2.....	31
2.4 Epidermal growth factor receptor (ErbB2).....	32
2.5 Computational analyses of ErbB2 and miR-17-3p.....	32
2.6 Elucidation of possible miR-17-3p sites between 44/576, 110/576, and 45/576 within the 3'UTR of ErbB2.....	35
2.7 The goals of this chapter are as follows.....	37

2.8 Methods and Materials.....	38
2.9 Results.....	41
2.10 Discussion.....	43
Chapter 3- The significance of hsa-miR-299-5p, and possibly hsa-miR-299-3p, contributing to prostate cancer progression.....	46
3.1 Introduction and previous work.....	47
3.2 Hsa-miR-299-5p.....	50
3.3 Osteopontin (OPN).....	50
3.4 Binding of miR-299-5p to the 3'UTR of OPN.....	51
3.5 Hsa-miR-299-3p.....	52
3.6 Experimental Aims.....	53
3.7 Methods and Materials.....	54
3.8 Results.....	62
3.9 Discussion.....	76
Chapter 4- Determine the expression levels of the five selected messenger RNA targets for miR-147b, an oncomiR.....	81
4.1 Introduction.....	82
4.2 A brief overview of the 5 mRNA genes.....	85
4.3 RNAhybrid analyses for miR-147b and its targets.....	87
4.4 The goals of this research.....	89
4.5 Methods and Materials.....	90
4.6 Results.....	92
4.7 Discussion.....	95
Chapter 5- A bioinformatics approach to identify microRNAs present in the body fluids of normal subjects.....	98
5.1 Introduction.....	99
5.2 Stability of miRNAs.....	100
5.3 Identifying miRNAs in body fluids.....	101
5.4 Aim of this research.....	102
5.5 The CLC genomics workbench.....	103
5.6 Methods to curate RNA-SEQ data in the CLC genomic workbench.....	104

5.7 Results.....	109
5.8 Discussion.....	126
Chapter 6- Conclusions.....	137
References.....	143
Vita.....	154

List of Figures

	Page
Figure 1-1: The TNM system of prostate cancer.....	6
Figure 1-2: MicroRNA biogenesis.....	12
Figure 1-3: Development of the M12 cell line from the P69 cell line.....	15
Figure 1-4: Previous results showed that miR-17-3p targets vimentin.....	17
Figure 1-5: Human panel analyses of miRNA expression in P69 vs. M12 cell lines.....	19
Figure 1-6: Comparing the expressions of miRNAs between the human panel and single-miR confirmatory analyses.....	22
Figure 1-7: Measuring expressions of oncomiRs and tumor suppressors in prostate cancer cell lines.....	24
Figure 1-8: Expressions of miRNAs observed in the laser capture microdissected tumor tissue samples.....	25
Figure 2-1: The 3'UTR of ErbB2 showing the three binding sites of miR-17-3p at positions 44, 110, and 451 extending to the common 3'-end at nucleotide 576.....	36
Figure 2-2: Luciferase assays of miR-17-3p's binding sites within ErbB2's 3'UTR constructs of 44/576, 110/576, and 451/576.....	42
Figure 3-i: Differential expressions observed between miR-299-3p and -5p for panel and single-miR analyses.....	47
Figure 3-ii: The expression levels of miR-299-3p and -5p observed for prostate cancer patients in benign and tumor samples.....	49
Figure 3-1: The expression levels of miR-299-3p, miR-299-5p, miR-375, and miR-141 in the M12 and M12+miR-299 cell lines.....	63
Figure 3-2A: The repression rates of 3'UTR OPN 9 and 3'UTR OPN 13 observed in the M12 cell line.....	67

Figure 3-2B: The repression rates of 3'UTR OPN 9 and 3'UTR OPN 13 observed in the M12+miR-299-5p cell line.....	68
Figure 3-3A-C: The breakdown of old sets of original sequences of 3'UTR OPN 9 and OPN 13 into new sets of original sequences and their mutants.....	70
Figure 3-4: The increased expression rates for the mutated OPNs in the M12+miR-299-5p cell line.....	72
Figure 3-5: Cell proliferation conducted on M12 and M12+miR-299-5p at 24, 48, and 72 hours show that M12 and M12+miR-299-5p grow nearly at the same rate.....	74
Figure 3-6 A-D: A-B Migration assay for early and late cell culture passage of M12 and M12+miR-299-5p cell lines. C-D: Invasion assay for early and late cell culture passage of M12 and M12+miR-299-5p cell lines	75
Figure 4-1: Luciferase assay results for transfection of the psiCHECK-2 vector containing the 3'UTR of 5 mRNA targets of miR-147b into the M12 cell line compared to the empty vector.....	93
Figure 5-1: The final PCR product and the location of each adapter sequence.....	105
Figure 5-2: Examples showing how mismatches and gap-costs are calculated based on the Smith Waterman algorithm.....	107
Figure 5-3 A-L: Common mature 5' and 3' miRNAs for each body fluid.....	125

List of Tables

	Page
Table 1-1: MiRNA selected for single-miR analyses.....	21
Table 2-1: The three potential miR-17-3p binding sites (at 5'-position 44, 110, and 451 all extending to 576) within the 3'UTR of ErbB2 from RNA-hybrid software.....	34
Table 3-1: The miR-299 binding within the 3'UTR OPN 9 (56nt) and 3'UTR OPN 13 (55nt)..	65
Table 3-2: Comparison of mfe between the new shortened sets of 3'UTR OPN 9 (25nt) and 3'UTR OPN 13 (26nt) and their mutants.....	71
Table 4-1: Sequence similarity between miR-147a, miR-147b, and miR-210.....	83
Table 4-2: Binding of miR-147b to its 5 mRNA targets.....	88
Table 4-3: Sequences of the 5 mRNA targets of miR-147b.....	91
Table 4-4: Results obtained from RNAhybrid software, and luciferase assay.....	94
Table 5-1A: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for venous blood.....	110
Table 5-1B: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for menstrual blood.....	110
Table 5-1C: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for vaginal fluid.....	111
Table 5-1D: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for semen.....	111
Table 5-1E: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for saliva.....	112
Table 5-1F: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for feces.....	112
Table 5-2: The percentage each body fluid was annotated against the miRBase database.....	113
Table 5-3: Compiled version of miRNAs found commonly in two or more body fluids.....	115
Table 5-4: Annotation results of each body fluid.....	116

Abbreviations

µg	microgram
µl	microliter
aatf	anti-apoptosis transcription factor
Ago	Argonaut
ALDH5A1	aldehyde dehydrogenase 5 Family, Member A1
AML	acute myeloid leukemia
AR	androgen receptor
BPH	benign prostatic hyperplasia
cDNA	complementary DNA
COL4A2	Collagen alpha-2(IV) chain
DGCR8	DiGeorge Syndrome Critical Region Gene 8
DNA	deoxyribonucleic acid
DU145	metastatic prostate cancer cell line
EDTA	Ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EMT	epithelial to mesenchymal transition
ErbB2	v-Erb-B2 Erythroblastic Leukemia Viral Oncogene Homolog 2
F6	poorly tumorigenic prostate cell line
FFPE	formalin-fixed paraffin embedded
G1	phase of cell cycle
G2/M	check point of mitotic cell cycle
H ₂ O	water
hsa	homo sapiens (human)
IER5	Immediate early response 5
LCM	laser-captured microdissection
LNA	locked-nucleic acid
LNC RNA	long non-coding RNA
LnCaP series	hormone-progressive prostate cancer cell lines
M12	metastatic prostate cancer cell line
M2182	moderately metastatic prostate cancer cell line

mcs	multiple cloning site
mfe	minimum free energy
miRNA, or miR	micro-RNA
mL	milliliter
mRNA	messenger RNA
NARF	Nucleic Acid Research Foundation
NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex,
NPC	nasopharyngeal carcinoma
OPN	Osteopontin or Secreted phosphoprotein 1
P69	benign epithelial prostate cell line
PBS	phosphate-buffered saline
PC3	metastatic prostate cancer cell line
PCR	polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
PSA	prostate-specific antigen
PTEN	Phosphatase and tensin homolog
QPCR	quantitative polymerase chain reaction
Rb	Retinoblastoma protein
RISC	RNA-Induced Silencing Complex
RNA	ribonucleic acid
RNU	Small Nucleolar RNA (also SNORD)
RPMA	Reverse-phase Protein Microarray
RPMI	Roswell Park Memorial Institute medium
SDHD	Succinate dehydrogenase complex subunit D.
SV40T	simian virus 40 large T antigen
UNIX	uniplexed information and computing system
UTR	untranslated region
UV	ultraviolet
VCU	Virginia Commonwealth University

Abstract

Identification of micro-RNAs and their messenger RNA targets in Prostate cancer and Biological fluids

by

Kanika Sharma, B.Sc.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Prostate cancer is the second most common cancer in the United States that affects men today. To better treat this disease accurate biomarkers and successful therapeutic treatments are needed. A novel approach to understand the mechanisms behind prostate cancer tumor formation lies in identifying dysregulated micro-RNAs (miRNAs), which are a class of small (18-24 nucleotides) non-coding RNAs that regulate gene expression posttranscriptionally by either inhibiting protein synthesis or signaling messenger-RNA for degradation. Multiple miRNAs were discovered in our highly tumorigenic and metastatic prostate cancer progression model M12 cell line compared to its weakly tumorigenic P69 parental cell line. Various analyses such as human panel analyses, single-miR analyses and patient tumor biopsy samples were analyzed to determine dysregulated miRNAs that contributed to the progression and metastasis of prostate cancer. Together with performing experiments to identify miRNAs, a *de novo* next generation sequencing approach was applied to identify miRNAs naturally present in biological fluids of normal and healthy subjects. Since, these miRNAs are highly dysregulated in many diseases, including cancer, they can act as

potential biomarkers or therapeutic targets to improve treatments for prostate cancer. Essential miRNAs studied for this research are miR-17-3p that is known to target the ErbB2 mRNA; miR-299-5p that directly targets osteopontin (OPN), and miR-147b which directly targets many mRNAs, such as COL4A2, ALDH5A1, NDUFA4, SDHD, and IER5. A wide range of miRNAs were identified in six biological fluids: venous blood, menstrual blood, vaginal fluid, semen, saliva, and feces. There were some miRNAs that were common to all 6 body fluids, and some miRNAs that literature suggested could potentially be biomarkers or normalizers for body fluid characterization.

Chapter 1:

Introduction, background, and foundation of miRNA research

1.1 Prostate cancer and its origin

Over the past several years, research in the field of cancer has elevated exponentially. Cancer is a heterogeneous disease that is a result of multiple gene mutations, aberrant gene expressions, and microRNA (miRNA) dysregulation [4, 5]. A host of mutational events in DNA, combine to dysregulate multiple signaling pathways. Bioinformatics analyses have elucidated a dozen more signaling pathways, which impinge tumorigenesis, and a range of mutations that differ even in similar tumors [1]. Furthermore, high-throughput sequencing data have revealed that there are many more point mutations, translocations, amplifications, and deletions that heavily contribute to cancer development [2, 3].

In the U.S. one of the most commonly diagnosed cancers, and a leading cause of cancer death in men is prostate cancer [6]. Prostate cancer develops in the prostate gland, which forms a part of the male reproductive system. The main function of this gland is to produce seminal fluid, which is alkaline in nature, and produced to protect and enrich the sperm, thereby enhancing its survival for an elongated period of time [7]. All prostate related diseases, both benign and malignant, stem from this walnut-shaped gland located at the base of a man's bladder. Normal prostate epithelium is composed of basal and luminal cells. In prostate cancer, the luminal and basal cell arrangement is disrupted and a loss of basal cells is observed. The expansion of luminal cells expressing prostate-specific antigen (PSA) is associated with the origin of human prostate cancer. Activation of the P13K pathway due to the loss of PTEN, is also frequently observed in human prostate cancer. Although luminal cells, commonly being the origin of prostate cancer, are used as therapeutic targets, now basal cells are showing potential as therapeutic targets, due to the presence of prostatic stem cells in the basal epithelium compartment [8].

A variety of genetic and epigenetic factors such as age, race, heredity, diet, sexual, and physical activity contribute to the development of prostate tumors. Age especially, is an important factor. Young and healthy men have a small prostate while older men, around the age of fifty, tend to have an enlarged prostate. The increase in size of the prostate is due to male hormones called androgens, such as testosterone [9]. The growth of the prostate can either be classified as benign or malignant. Benign cases are often referred to as benign prostatic hyperplasia (BPH), where the prostate enlarges and presses on the urethra, resulting in irregular flow of urine. Since, these are benign cases, their overgrowth does not invade other tissues or organs. In malignant cases, the cells invade and damage nearby tissues, eventually leading to metastasis to bone and brain [10].

1.2 Common occurrences of prostate cancer

According to the American Cancer Society, in the year of 2014, a total of 233,000 new prostate cancer cases are being estimated while 29,480 deaths are being predicted. It is a disease that affects approximately 1 out of 6 men in their lifetime [11]. It has been estimated that every 2.3 minutes, a new case of prostate cancer occurs, while every 16 minutes a man dies from prostate cancer. In fact more than 2.5 million American men are currently living with prostate cancer. When compared to other cancers it has been estimated that a man is 35% more likely to be diagnosed with prostate cancer than a woman is to be diagnosed with breast cancer. African American men are 56% more susceptible to prostate cancer and 2.5 times more likely to die from this disease than the Caucasian men. Men, aged 65 or above, with a family history of prostate cancer are twice as likely to develop this disease [10].

1.3 Screening of prostate cancer

When considering the risks of prostate cancer, several factors are taken into account such as, the grade of cancer, (Gleason score), the PSA level, and the stage of cancer- basically how large is the tumor is and has it metastasized [12].

The Gleason Score enables physicians to grade the prostate cancer. It is considered as one of the most reliable methods for estimating the aggressiveness of prostate cancer. A Gleason score is the sum of two scores, where the primary Gleason grade is added to the secondary Gleason grade to produce a score. These grades range from 1-5; where a grade of 1, basically resembles the normal prostate whereas a grade of 5 is considered an extreme case of cancer. The Gleason sum ranges from 2-10 where an extreme case of prostate cancer is denoted by the score of 10 [12].

The PSA is a protein produced by both normal and prostate cancer cells. In order to detect signs of prostate cancer, the PSA test measures the level of PSA in a man's blood. Results are usually reported per milliliter of blood. PSA levels under 4 ng/ml are considered healthy for men and as the PSA level rises the chances of having prostate cancer also rises. Usually, a PSA level less than 10 is considered a low risk, while 10-20 is considered an intermediate risk, and >20 is considered high risk [13]. Identifying PSA levels maybe a good initial step for detecting the occurrences of prostate cancer, but it is not very specific. Sometimes men with even a high PSA level do not have cancer and this is because PSA levels can rise due to variety of benign conditions such as inflammation of the prostate, BPH, or prostatitis of prostate cancer. Therefore, PSA testing alone is not sufficient to detect prostate cancer; hence a digital rectal exam (DRE) should be used in combination with PSA monitoring for diagnosing the disease [13].

The stage of cancer is determined based on the size of the tumor and how much it has metastasized. Typically, these stages are expressed as the T, N, and M score, and are collectively referred to as the TNM system for staging prostate cancer. T plus a number (0-4) is used to determine the location of tumor, where a score of T0/T1 indicates a tiny to no tumor and is only picked up by PSA testing or sometimes not picked up by any sort of imaging. A Score of T4 indicates the tumor is fixed or it is beginning to invade the nearby tissues or organs, other than the seminal vesicles. N stands for the lymph nodes and it ranges from 0-1, where N0 indicates that cancer has not spread to the lymph nodes, and N1 indicates it has. M in the TNM system indicates the metastasis of tumor to other parts of the body such as the lungs or bones. It ranges from 0-1, where M0 indicates the disease has not spread, while M1 indicates it has begun to spread beyond the lymph nodes (Figure 1-1) [10, 13, 14].

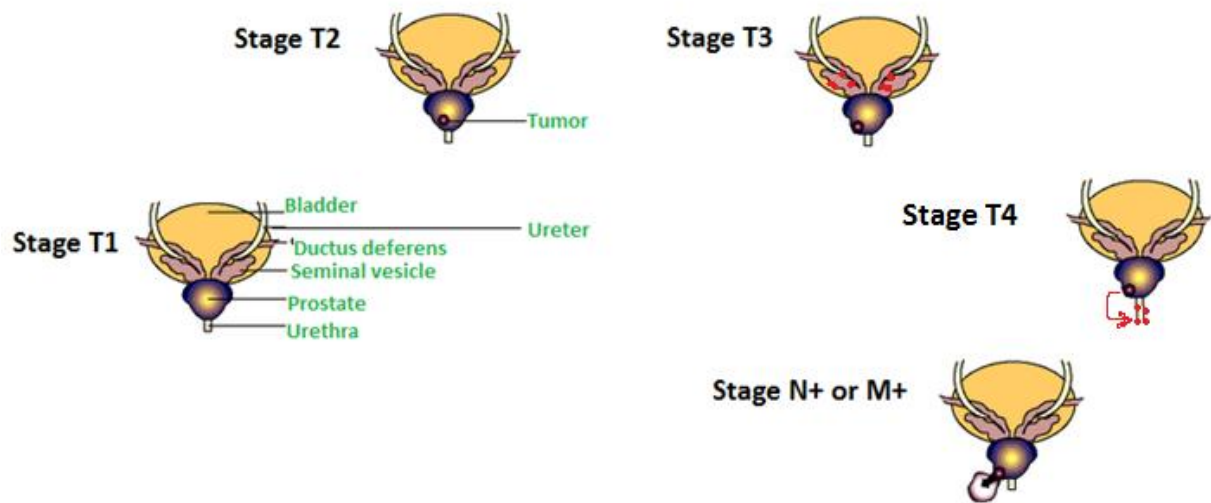


Figure 1-1: The TNM system of prostate cancer. Stage T1 indicates clinically the tumor was not found by examination or by the use of imaging such as a transrectal ultrasound. Stage T2 indicates a tumor found confined to the prostate, in either one or both lobes by digital rectal exam (DRE) or transrectal ultrasound. Stage T3 indicates the tumor has extended through the capsule surrounding the prostate and may have begun to spread into the seminal vesicles. Stage T4 indicates the tumor has metastasized to nearby tissues or organs, other than seminal vesicles, such as the urethral sphincter. Stage N+ indicates that the lymph nodes show incidence of prostate cancer, while stage M+ indicates that cancer has slowly begun to spread to distant organs [12-14].

1.4 Treatments for prostate cancer

Prognosis of prostate cancer depends on whether the disease is non-metastatic or metastatic. For non-metastatic disease age, clinical stage, Gleason score, PSA level, and co-morbidity are important factors. If prostate cancer is discovered at an early stage, the majority of the times it can be cured. For metastatic cases age, site of metastases, weight loss, hemoglobin level and androgen dependence are important factors. Once the tumor extends beyond the prostate gland, it becomes fatal and survival rates drop, hence currently for metastatic cases, no curative therapy exists [15].

Treatments are based on the type of cancer and the aforementioned factors, including life expectancy and personal preference. For an unfit elderly man with non-metastatic prostate cancer such as T1 or T2 tumors the patient is more likely to die due to other reasons than the cancer itself. For patients whose life expectancy is less than 10 years, their personal preference may include no treatment. Otherwise, radiotherapy or brachytherapy are often suggested for the elderly. For a healthy elderly or a younger adult with non-metastatic prostate cancer, the life expectancy is often more than 10 years. In this case often radical prostatectomy, external beam radiotherapy, or brachytherapy are suggested in an attempt to eradicate tumor. For cases where prostate cancer has locally invaded the seminal vesicles, radiotherapy or hormonal therapy is recommended. Patients who have been newly diagnosed with metastatic prostate cancer, together with radical prostatectomy undergo androgen deprivation and hormonal therapy. The optimal form of hormonal therapy is orchiectomy, which is removal of one or both testicles, gets rid of almost 90% of circulating androgens and the remaining 10% is produced by the adrenal glands [15].

Generally, within a few years of initial treatment, the tumor re-occurs in 15-30% of prostate cancer patients, eventually leading them to androgen deprivation therapy. This therapy has a higher success rate than other therapies, usually 70-80% of patients respond to it positively; however, it is only effective for patients with androgen-dependent prostate cancer. In a normal prostate cell and in the androgen-dependent prostate tumors, testosterone binds to the androgen receptor, which instigates the production of androgen and promotes cell growth. As tumor progresses to later stages, it becomes androgen-independent and fails to respond to the androgen deprivation therapy [17].

There is an ongoing research on how the androgen receptors are able to promote growth in the absence of testosterone. Studies have shown that 50% of androgen-independent tumors contain point mutations in the androgen receptor gene. Hence, there are speculations, that these point mutations play a role in tumorigenesis [17, 18].

Although there is chemotherapy, differential therapy, and numerous forms of hormonal therapy to alleviate the progression of prostate cancer, these are temporary and provide relief for only a short period of time. Serving to either slow down progression, or if the prostate cancer is in its early stage, attempt to eradicate cancer cells. To overcome the shortcomings of these therapies further research on the growth of cancer, the nature of the tumor and the mechanism behind its formation, must be elucidated.

1.5 Identifying biomarkers of prostate cancer

Prostate cancer is highly heterogeneous in nature and multiple gene mutations play an important role in tumor progression. Lately, studies have revealed that apart from essentially identifying genes that promote tumor progression, protein and RNA also contribute to tumor progression and can potentially serve as biomarkers for this disease [16]. In fact, much of the research is currently being focused on microRNAs (miRNAs, miRs), which are small non-coding RNAs that influence the expression of a gene post-transcriptionally, either by inhibiting translation or targeting the transcribed messengerRNA (mRNA) for degradation.

1.6 Micro-RNAs and their role in oncogenesis

Micro-RNAs are single stranded non-coding RNAs, 18-24 nucleotide (nt) in length, that post-transcriptionally regulate gene expression either by inhibiting translation or marking a mRNA for degradation. It has been estimated that miRNAs modulate almost 30-60% of protein coding genes [19]. Along with that they are also involved in the regulation of a variety of biological processes such as cell cycle differentiation, development, metabolism, cell proliferation, apoptosis, response to stress, as well as human diseases including diabetes, immunological or neurodegenerative disorders, and cancer [20, 21].

MiRNAs were first discovered in 1993 in the nematode *Caenorhabditis elegans* [22]. Lee *et al.* discovered that the gene *lin-4*, which was an essential gene for *C.elegans* post-embryonic development, did not code for any protein, but was instead transcribed into a 22-nucleotide RNA molecule, which was able to inhibit the expression of the *lin-14* mRNA by directly targeting its 3' untranslated region (3'-UTR) [23]. Another case of RNA discovery was of the *let-7* small RNA that was also involved in the developmental timing of *C. elegans*. When it was sequenced, it was

found to be highly conserved in a variety of organisms, which suggested that these small RNAs possibly play an important role in regulating gene expression outside of *C. elegans*. Soon, many miRNAs were discovered in worms, flies and mammalian genomes [24].

The role of miRNAs is critical in tumor development as they regulate the cellular processes of oncogenes or tumor suppressor genes [26]. They are frequently located in cancer-related genomic regions such as regions of amplification, fragile sites, or common break point regions on or near oncogenes or tumor suppressor genes. Oncogenes are consistently observed to be up-regulated in tumor versus normal tissue, whereas tumor suppressor genes are consistently down-regulated in tumor versus normal tissue. In a given cancer miRNAs are either known to act as an oncomiR, where it gets overexpressed, or tumor suppressors, where its expression is lost. Due to miRNA dysregulation of miRNAs, increased expression of oncomiRs contributes to oncogenesis, while loss of tumor suppressors can also result in increased oncogenesis. Whether tumor suppressors are lost or oncomiRs overexpressed, the end result is cell growth, proliferation, tumor progression and ultimately metastasis [19, 26-27].

1.7 Biogenesis of miRNAs

MiRNA genes are evolutionarily conserved across species and are potentially located either within introns or exons of protein coding genes or in the intergenic regions. These small non-coding RNAs are transcribed in the nucleus by RNA polymerase II or sometimes RNA polymerase III into larger RNA precursors as primary-miRNAs (pri-miRNAs), which are often several kilobases long. These pri-miRNAs are then capped (MGpppG) and polyadenalated (AAAAA) [19] and further processed by the RNase III enzyme Drosha and its co-factor DGCR8/Pasha, to generate an ~70 nucleotides long premature-miRNA (pre-miRNA). This pre-miRNA folds into a

stem-loop structure and gets exported from the nucleus into cytoplasm by a GTP-dependent exportin 5 transporter. Pre-miRNA is further processed by an RNase III enzyme called Dicer, which together with its cofactor TRBP snips off the loop and releases a double stranded RNA duplex approximately 18-24 nts long. This double stranded RNA is then incorporated into the miRNA-associated multiprotein RNA-induced silencing complex (RISC) that includes the Argonaute proteins to help select the single stranded mature miRNA. Together with the mature miRNA bound within the RISC complex, mRNAs are targeted either for inhibition by blocking the translation of mRNA into proteins or for mRNA degradation. The first 5-7 bases on the 5' end, also known as the “seed region” of the miRNA, are known to be required for binding to the complementary sequences of their target mRNA (Figure 1-2) [19, 28].

The inhibition or destruction of an mRNA depends on the amount of sequence match between the miRNA and mRNA. When the miRNA matches perfectly with the mRNA, the result is degradation of the mRNA by cleavage between base 10 and 11 of the miRNA as well as the complementary region of the mRNA target. However, when only a few nucleotides match the result is inhibition of mRNA translation. The miRNA's ability to bind to its target mRNA either perfectly or imperfectly enables it to act as a final modulator of protein expression. Alterations in normal expression patterns can contribute to cancer. Since cancer can result from the improper expression of a disordered genome where abnormal cells can grow uncontrollably, and metastasize to other parts of the body, miRNAs, are known to play a critical role in carcinogenesis and tumor progression. Bioinformatics data have shown that a microRNA can target many mRNAs as well as one mRNA can be targeted by many miRNAs [29].

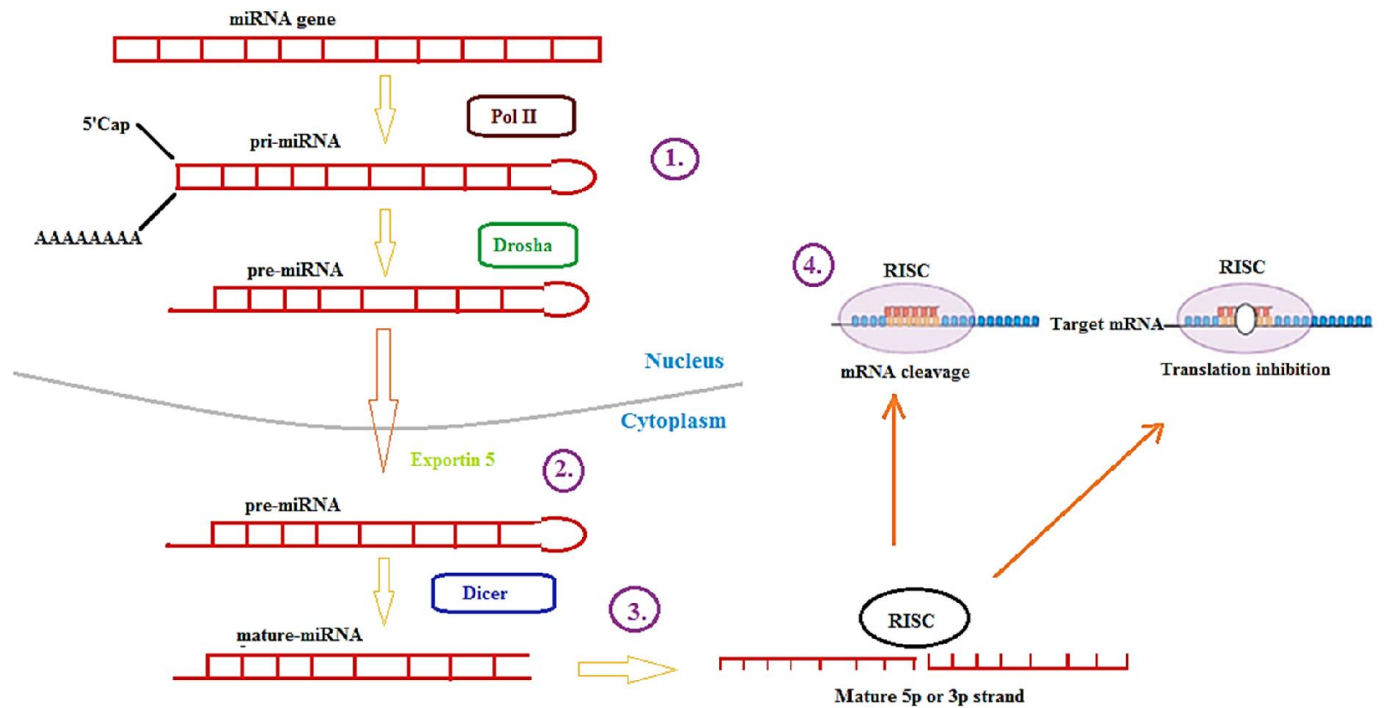


Figure 1-2: MicroRNA biogenesis. (1): miRNA gene is being transcribed by RNA polymerase II into pri-miRNA of variable size, which is further cleaved by the RNase III called Drosha with DGCR8, to produce a hairpin precursor form called pre-miRNA, which is approximately 70 nt long. (2): Pre-miRNA is exported from the nucleus into the cytoplasm by exportin 5 and further snipped off at the loop by another RNase III enzyme the Dicer. (3): TRBP complex generating a miRNA duplex of 18-24 nt in length. One strand is selected by the multiprotein RNA-induced silencing complex (RISC) by an unknown mechanism. (4): The mature miRNA depending on the complementary match to its mRNA, leads RISC to either inhibit mRNA's translation or completely degrade the mRNA [19].

1.8 Nomenclature of miRNAs:

For a given miRNA named “hsa-miR-147b”; the “hsa” refers to the species, in this case *Homo sapiens* miRNA, “miR” refers to the miRNA precursor, “147” refers to the 147th family that was named, and “147b” refers to its relativity to other miRNAs within the 147th family. For example, there must be at least one other family member as hsa-mir-147a. Other small letters such as from c-h refer to other family members, which all share the same "seed sequence" and differ by at most 1-2 nucleotides outside of the crucial seed region, hence they are included in the same miRNA family. Family members exhibit tissue-specific expression so various tissues/cell-types will differ in which family members are expressed. If a given name is “hsa-mir-299-3p” or “hsa-mir-299-5p”, the same naming process applies till the 299th part; however, the 3p and 5p refers to the mature miRNA strand from either 3’ and 5’ end of the precursor, which means that hsa-mir-299-3p was produced from the 3’ end of the mature miRNA, and hsa-mir-299-5p was produced from the 5’ end of the mature miRNA [25].

1.9 Research objectives, previous works, and models used:

The objective of our laboratory’s research is to identify key miRNAs, the dysregulation of which can contribute to prostate tumor progression. Dysregulated miRNAs that are involved in tumorigenesis can cause a variety of expression differences contributing to cancer development. *In-vivo*, *in-vitro*, and *in-silico* methods have been applied for miRNA identification purposes. Moreover, we seek the identification of miRNAs that can be used as biomarkers for prostate cancer identification, staging, and treatment.

1.9.1 Prostate cancer progression model M12 cell line:

To investigate the progression of prostate cancer and to study the overall cellular behavior of a unique prostate cancer cell line the P69 cell line was developed. This cell line was developed from human non-neoplastic prostate epithelial cells isolated from a 63 year old African American male and immortalized with the simian virus 40 (SV40) large T antigen gene that inactivates the cellular break, the retinoblastoma protein (Rb). The P69s are poorly tumorigenic and non-metastatic. A highly tumorigenic cell line was developed from the P69s by subcutaneously injecting in male athymic nude mice. Out of the eighteen mice that were injected only two mice developed tumors. After two successive rounds of such injections, a variant cell line, called the M2182 cell line, was developed. At each successive round of injections the tumors grew bigger and faster. On the third round of injections, another new cell line, called the M12 cell line was derived. Upon subcutaneous, intraperitoneal, and intraprostatic injections into the prostate it was found that the M12 cell line was not only highly tumorigenic, but it easily metastasized to multiple secondary sites in the animal such as the diaphragm, lung, and mesentery. Within the prostate field this is unique as all other prostate cell lines are unable on their own to escape from the confines of the prostate making the M12 cell line highly useful. In conclusion the P69, M2182, and the M12 cell lines are tumorigenic models used to study prostate cancer (Figure 1-3) [30].

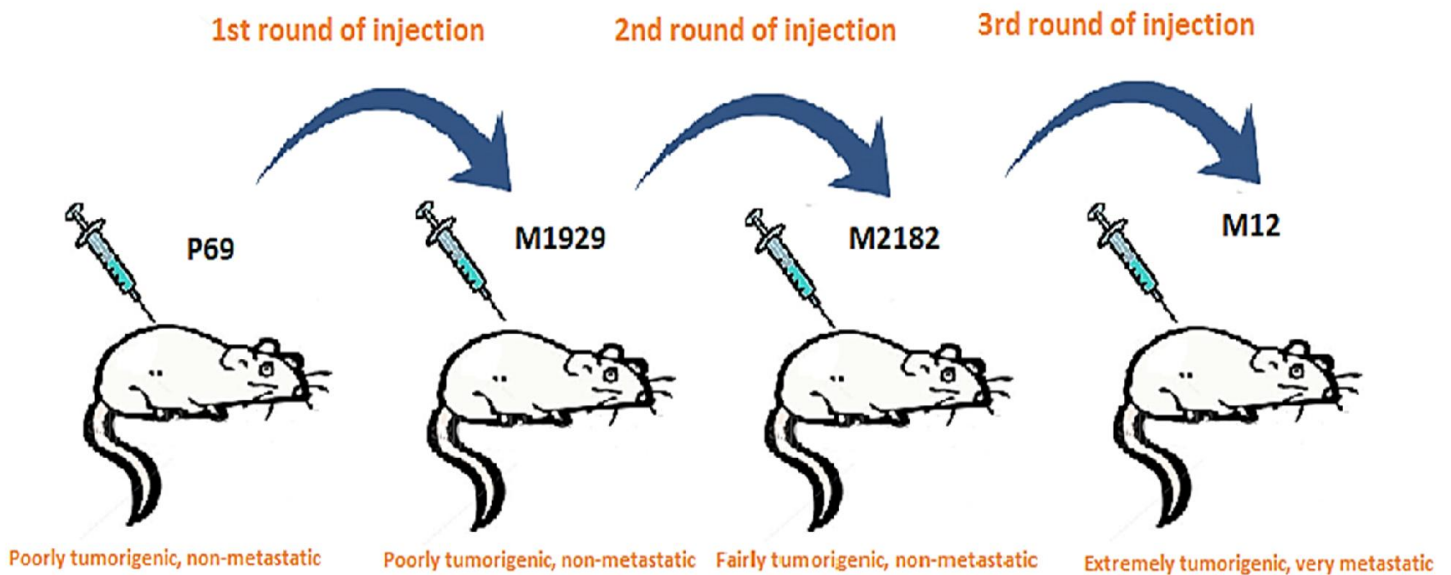


Figure 1-3: Development of the M12 cell line from the P69 cell line: Human prostate epithelial cells were immortalized with SV40 large T antigen and injected into athymic nude mouse. A series of injections were carried out *in-vivo* for selection process. During the first round of injection, the M1929 cell line developed, which was poorly tumorigenic and non-metastatic. From the second round of injection the M2182 cell line developed, which was fairly tumorigenic, but still not metastatic. Finally, from the third round of injections the M12 cell line was obtained which was highly tumorigenic and very metastatic when injected orthotopically.

1.9.2 Other prostate cancer cell lines:

Three other sets of tissue culture cell lines used by researchers and our laboratory, to study the overall progression of prostate cancer are as follows: the Dunning-145 (DU-145), LNCaP, and PC3 cell lines. The first prostate cancer cell line to be developed was DU-145 cell line. This cell line was derived from a metastatic tumor excised from the brain of a 69 year old white male with prostate cancer. The LNCaP cell line was developed from a biopsy of a lymph node metastases from a 50 year old white male with prostate cancer. There are more than 60 sublines for this cell line, and the parental cell line expresses both the normal androgen receptor (AR) mRNA and protein. The PC3 cell line was developed from a lumbar vertebra metastasis from a 62 year old white male with prostate cancer. There are a total of 11 sublines for this cell line, some express the AR gene, and some do not [31].

1.9.3 Identification of dysregulated miRNAs through the P69-M12 model cell lines

Previous work performed in the laboratory revealed that miR-17-3p, a miRNA found on chromosome 19, to be down-regulated in the M12 cell line relative to the P69 cell line. This loss contributed to the expression of vimentin mRNA, which encodes for an intermediate filament protein up-regulated in the M12 compared to the P69 cell line. The high expression of vimentin was further reversed by stably transforming the M12 cell line with a plasmid vector expressing miR-17-3p, known as the M12+miR-17-3p cell line. The reduced expression of vimentin in the M12+miR-17-3p cell line indicated that miR-17-3p may potentially be binding to the 3'UTR of vimentin mRNA (Figure 1-4) [32]. Similarly, miR-125b was identified as a tumor suppressor due to a loss of expression observed in the M12 versus P69 cell line. A reduction in migration and invasiveness was observed when miR-125b was restored in the M12 cell line through stable

transformation, to generate the M12+miR-125b cell line. The mRNAs targets for miR-125b are ErbB2 & 3, which are found to be up-regulated in cancer [33].

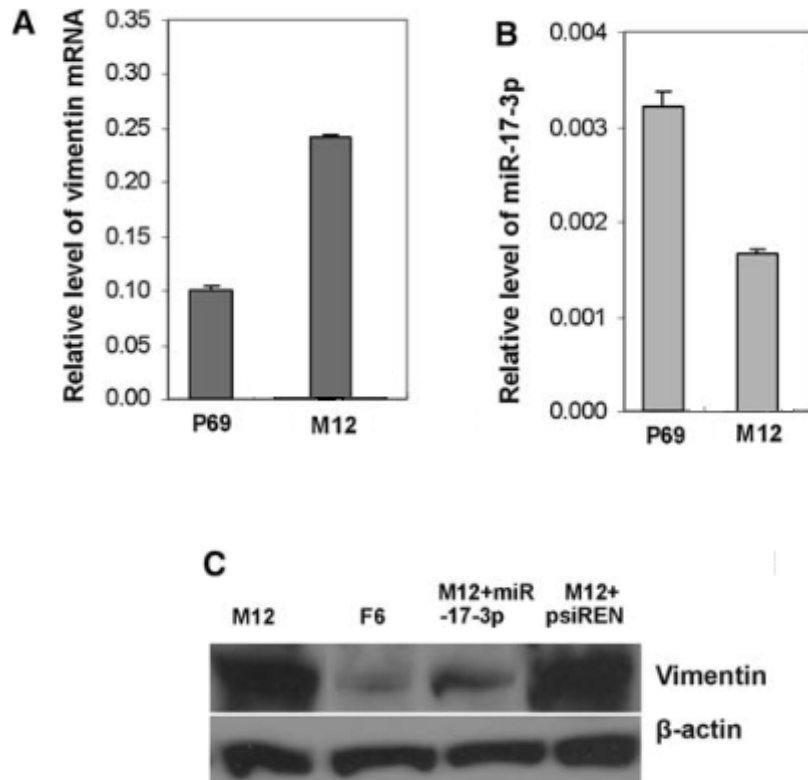


Figure 1-4: Previous results showed that miR-17-3p targets vimentin. A- A gain of expression was observed for vimentin mRNA in the M12 cell line. B- Reduced expression of miR-17-3p observed in the M12 cell line versus the P69 cell line. C- Stable restoration of miR-17-3p in the M12 cell line (M12+miR-17-3p) resulted in a substantial decrease in vimentin levels. (Adapted from Zhang et al. [32]).

1.10 Selection of miRNAs through screening processes

The reduced expression of vimentin mRNA observed due to targeting by miR-17-3p, sparked an interest in miRNAs in our laboratory. Since then we have continued to search for additional miRNAs that could serve as oncomiRs or tumor suppressors driving prostate cancer. Together with this goal, we are also interested in identifying potential mRNA targets for these miRNAs, the differential expression of which contributes to the cellular behavior and mechanisms of prostate cancer progression in our model M12 cell line.

In order to identify important miRNAs in the progression model M12 cell line, a series of analyses were performed as follows; (1) microarray panels consisting of a total of 736 miRNAs were analyzed followed by, (2) Single-miR confirmatory analyses followed by, (3) miRNA expression analyses in additional prostate cancer cell lines, and finally, (4) analyzing dysregulated miRNAs through laser-capture microdissection (LCM) of benign, stromal, and tumor tissue from human patient biopsies [34].

The entire credit for conducting analyses to select important miRNAs in prostate cancer goes to Drs. Sarah Seashols, William Budd, and Zendra Zehner. The order in which the names are listed by no means indicate the level of effort put into this project.

(1): Of a human panel of 736 miRNAs, the expression levels of miRNAs dysregulated in the P69 versus M12 cell line were determined via Reverse Transcriptase-quantitative Polymerase Chain Reaction (RT-qPCR). This analyses determined miRNAs that were overexpressed (OncomiRs) or lost (Tumor suppressors) between the progression of the "normal" P69 to its highly tumorigenic/metastatic variant, M12.

Results: Out of the 736 miRNAs analyzed, 231 miRNAs were found to be oncomiRs exhibiting a ≥ 2 fold increase in the M12 vs. P69 cell lines. Conversely, 150 miRNAs exhibited a ≤ 0.5 fold decrease and were labeled as potential tumor suppressors. The remaining 355 miRNAs exhibited no significant change in expression (Figure 1-5).

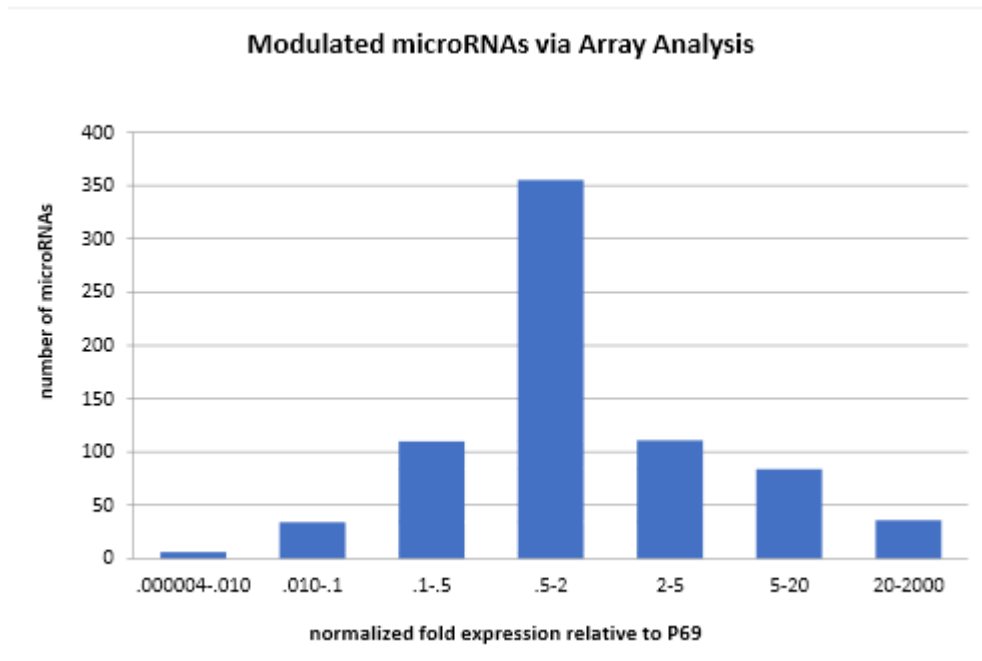


Figure 1-5 Human panel analyses of miRNA expression in P69 vs. M12 cell lines. Two human panel analyses were performed. Panel 1 was performed using 20 and 50 ng of RNA. Panel II was performed using 50ng of RNA. The global mean was used to normalize the data and a fold difference was calculated. Fold expressions greater than 1 indicated high expressions of miRNA and lower than 1 indicated low expressions of miRNA, in the M12 cell line relative to the P69 cell line. (Adapted from Seashols [Dissertation] [34])

MiRNAs chosen for confirmatory single miR analyses were not selected based not only on fold differences observed between the comparison of P69 to M12, but literature searches and bioinformatic tools were also incorporated to further confirm the selection. Furthermore, the miRNAs selected were a collection of both oncomiRs and tumor suppressors that showed relevance in prostate cancer.

(2): Single-miR analyses, of selected miRNAs as seen in Table 1-1, were conducted through RT-qPCR in RNA extracted from the M12 and P69 cell lines to confirm those trends exhibited in the human panel analyses.

Results: Most of the miRNAs were confirmed, but not all. MiRs-622, -221, and -299-3p were drastically different from the human panel analyses. The differences between panel analyses and single-miR analyses may be due to normalization process. The human panel data was normalized to the global mean whereas the single-miR data was normalized to RNU48. Although minor differences are anticipated due to differences in normalizations, but major differences should not account to different normalizations (Figure 1-6).

Table 1-1: MiRNA selected for single-miR analyses.

The highlighted miRNAs were selected for further studies in this thesis (MiRNAs and values for expression difference values adapted from Seashols [Dissertation] [34]).

MiRNA	Expression difference (M12 vs P69)
9	6.68
144	0.374
133a	37.91
133b	38.93
146a	0.459
147b	201.4
199a-3p	6.43
221	0.783
299-3p	11.37
299-5p	0.02
147	27.9
488	0.526
622	147.7
127-3p	0.002
127-5p	0.344

Expression differences between panel and single-miR analysis								
	133a	133b	9	622	199a-3p	147b	147	
Panel	37.9	38.9	6.7	147.7	6.43	201.4	27.9	
Single-miR	6.7	2.5	43.3	0.374	1.514	16.9	0.138	
	146a	488	221	127-3p	127-5p	299-3p	299-5p	144
Panel	0.459	0.526	0.783	0.002	0.344	11.4	0.02	0.374
Single-miR	0.021	0.039	1.39	0.003	0.164	0.035	---	---

Figure 1-6: Comparing the expressions of miRNAs between the human panel and single-miR confirmatory analyses. The panel analyses was normalized against the global mean and single-miR analyses was normalized against RNU48. Fold expression values of M12 cell line was set relative to the P69 cell line. Each sample was analyzed in triplicate. The ones outlined in red are miRNAs being studied for this thesis. The expression of miR-147b was consistent between the panel and single-miR analyses. MiR-299-3p was overexpressed in the panel analyses, but lost expression in the single-miR analyses. Loss of expression was observed for miR-299-5p in the panel analyses, but a value for single-miR analyses was not obtained due to a failed normalization expression value for the normalization standard to RNU48 (Adapted from Seashols [Dissertation] [34]).

(3): For additional confirmation, miRNAs that act as oncomiRs or tumor suppressors in both human panel analyses and single-miR analyses, were further evaluated in additional prostate cell lines, as following; PC3; DU-145 which is moderately metastatic; F6, a poorly tumorigenic, non-metastatic prostate cell line, obtained after re-introducing chromosome 19 into the M12 cell line; M2182, a moderately tumorigenic, non-metastatic cell line derived after second round of injections into the mouse, before development of the M12 cell line (Figure 1-7). Such comparisons permit the identification of miRNAs that behave identically in other relevant prostate cell lines, plus confirms which miRNA might be the most responsible for driving prostate tumorigenesis.

(4): Solid tumors like prostate are highly heterogeneous both in cell-type and stage of tumor development. In one tumor it is possible to find various stages of prostate cancer cells. These prostate cancer stages can be scored by the pathologist with a Gleason score ranging from 3-7 with an increase in number indicative of a higher staged cancer. Laser-capture microdissection (LCM) was used to extract RNA from formalin-fixed paraffin embedded (FFPE) or frozen tissue retrieved from prostatectomies to confirm via RT-PCR which miRNAs selected from the above analyses are truly dysregulated as normal glandular epithelial progresses to highly tumorigenic/metastatic cancer.

Results: Although the expression of some miRNAs varied from the results obtained above with human prostate cell lines, the expression of a few miRNAs remained consistent within the limited number (5) of patient biopsy samples available for LCM analysis. MiR-9 and miR-147b constantly came out to be as oncomiRs, whereas miR-299-3p, miR-199a-3p were tagged as tumor suppressors (Figure 1-8).

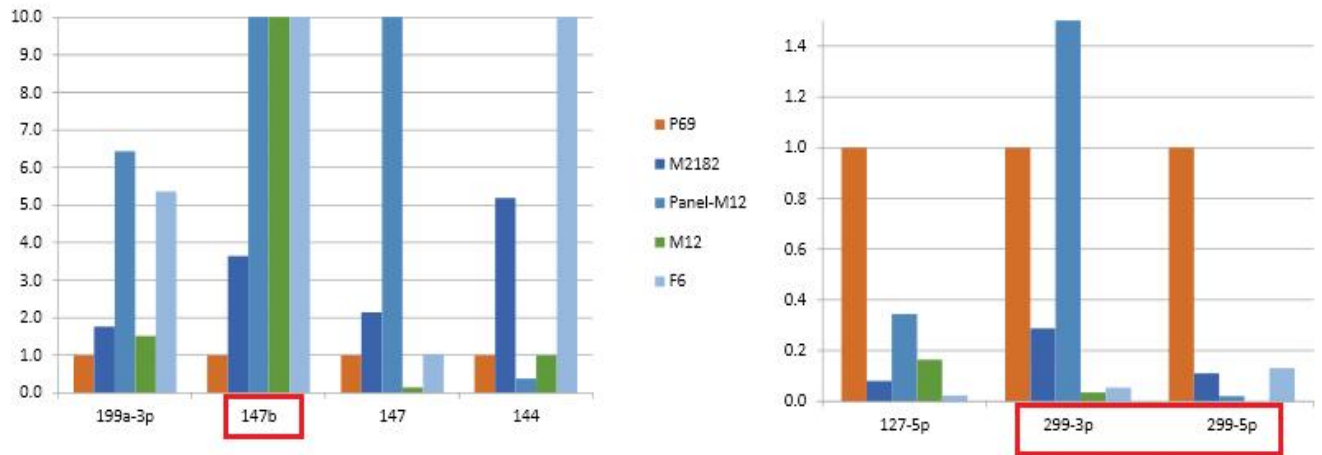


Figure 1-7: Measuring expressions of oncomiRs, and tumor suppressors in prostate cancer cell lines. MiRNAs levels were normalized to RNU48 and expression fold values of M12 are relative to the P69 cell line. Each sample was analyzed in triplicates. Those miRNAs outlined in red are being further studied in this thesis. MiR-147b is consistent in its expression, hence can be classified as an oncomiR. Reduced expression was observed for miR-299-3p in the M12, and F6 cell lines. No expression value was reported for miR-299-5p in the M12 cell line due to normalization discrepancy in RNU48. However, reduced expression of this miRNA was observed in other cell lines (Adapted from Seashols [Dissertation] [34]).

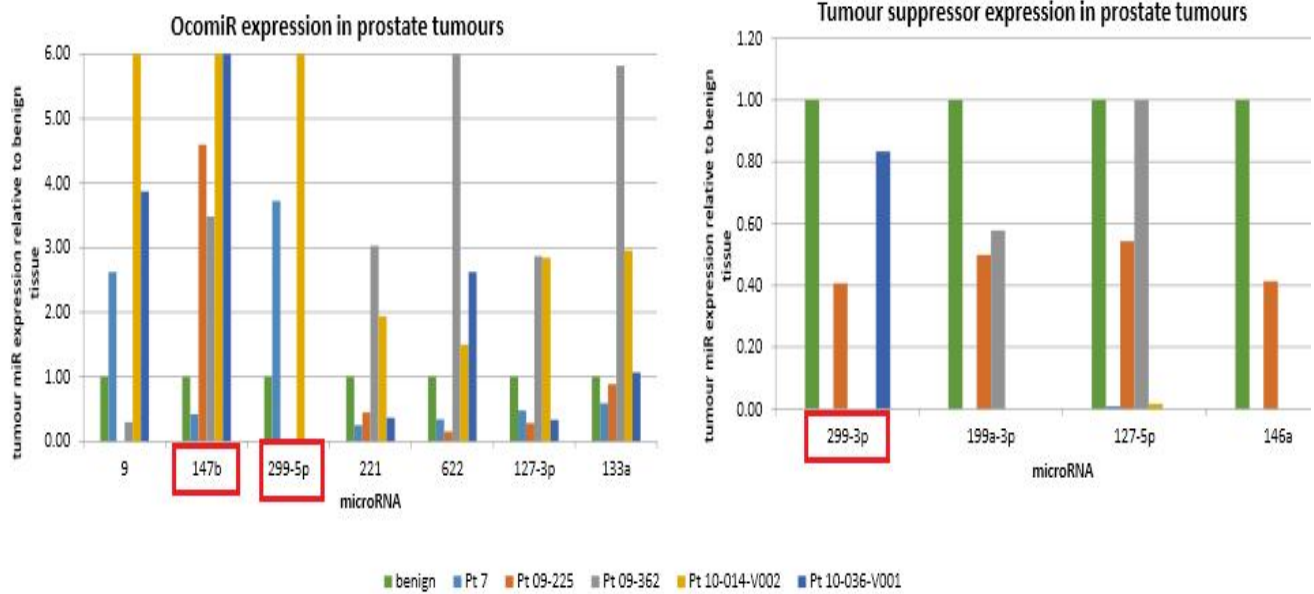


Figure 1-8: Expressions of miRNAs observed in the laser capture microdissected tumor tissue samples. MiRNA levels were normalized to RNU48 and fold expression values were relative to the benign tissue of the same patient. MiR-147b remained consistent in its expression acting as an oncomiR. MiR-299-3p has remained consistent in its expression acting as a tumor suppressor. MiR-299-5p is also observed as an oncomiR in the patient tumor samples. (Adapted from Seashols [Dissertation] [34])

1.11 The purpose of this thesis:

Identification of key oncomiRs or tumor suppressor miRNAs and their mRNA targets in prostate cancer through in-vitro and in-silico approaches.

The majority of work in the Zehner laboratory is devoted to identifying potential miRNAs that play an essential role in the progression of prostate cancer, by disrupting the cellular behavior to enhance proliferation, migration and invasiveness. Each chapter in this thesis presents an evaluation of miRNAs that were either highly dysregulated from cell lines to patient tumor biopsy samples, or showed great potential as key regulators of prostate cancer, or miRNAs that showed a direct connection between prostate cancer progression and the targets they prefer to bind. One of the chapters also presents an *in-silico* evaluation of miRNAs observed in the body fluids of normal and healthy people.

Chapter 2:

The work in Chapter 2 consists of determining which of the three potential miR-17-3p binding sites predicted in the 3'UTR of ErbB2, is/are most valuable for inhibiting the expression of ErbB2.

Chapter 3:

The work in Chapter 3 consists of determining potential mRNA targets for miR-299-5p. This miRNA was shown to be highly dysregulated in miRNA screen analyses and in the analyses of patient biopsy samples. Due to high variation observed in its expression, it became important to investigate the effects of miR-299-5p in prostate cancer.

Chapter 4:

The work in Chapter 4 consists of finding relevant mRNA targets for miR-147b, an oncomiR. MiR-147b has stayed consistent in its expression from cell line analyses to patient tumor biopsy sample analyses. However, scarce research has been performed on this miRNA. Due to its consistent nature as an oncomiR, we propose that this miRNA is essential for prostate cancer and finding its mRNA target(s) is important.

Chapter 5:

An *in-silico* approach was applied to identify conserved or abundant miRNAs in body fluids obtained from normal healthy people. The idea behind this research is that miRNA screen analyses are limited in their approach to identify all miRNAs present in prostate cancer tumor, however next generation sequencing gets rid of this limitation and can detect all miRNAs present in the tumor. We initiated the body fluids project with normal and healthy body fluids because, it is essential to first determine which miRNAs are prominent in normal body fluids and then determine/compare which miRNAs were either lost or gained in body fluids from patients.

1.11.1: In conclusion of this thesis (Chapter 6):

The work performed for this thesis culminates with an overall analyses that miRNAs are diverse in their nature, and depending on each cell type or tissue they are either expressed as oncomiRs or as tumor suppressors. This was the case in prostate cancer as well, where out of the 3 important miRNAs analyzed, two of them remained consistent in their expressions (miR-17-3p and miR-147b) and one differed (miR-299-5p). The *in-silico* approach helped determine miRNAs that were prominent not just in one body fluid but other body fluids as well. This indicates that

some of these miRNAs can serve as potential biomarkers for prostate cancer and help alleviate this disease by utilizing these miRNAs as potential therapeutic targets.

Chapter 2:

Elucidation and confirmation of miR17-3p binding sites within 3'UTR of ErbB2

2.1 History of miR-17-3p and its target vimentin:

As previously described, the Zehner laboratory works with a SV40T, immortalized human prostate cancer cell line progression model known as M12. This highly metastatic and tumorigenic cell line was developed from a parental cell line, P69, which is poorly tumorigenic and derived from normal prostatic tissue. Identifying potential miRNAs and their mRNA targets in the related M12 and P69 cell lines will aid in determining which miRNAs are required for the progression of prostate tissue from benign to tumor.

Vimentin is the type of intermediate filament protein (IFP) found in normal mesenchymal cells, where it plays an important role in maintaining cellular integrity and provides resistance against stress. It has been shown that vimentin levels are increased in epithelial cancers such as prostate cancer, gastrointestinal tumors, breast cancer, malignant melanoma, lung cancer and other types of cancers [42, 43]. Vimentin is an important marker for the epithelial-mesenchymal transition (EMT), which is known to be associated with tumorigenesis for a variety of different cancers [35]. Additionally, vimentin affects cell motility by enhancing metastasis in highly tumorigenic cell lines. One study showed that vimentin was more likely to be expressed in highly metastatic and tumorigenic cell lines than in slightly tumorigenic or non-metastatic cell lines [36]. Similarly, previous work done in the Zehner laboratory showed that vimentin mRNA was highly expressed in the highly tumorigenic and metastatic M12 cell line, but expressed less in the poorly tumorigenic and non-metastatic P69 cell line [34]. In order to control the high level of vimentin in the M12 cell line, miR-17-3p was identified as one miRNA capable of binding to vimentin's 3'UTR. The parental P69 cell line contained miR-17-3p and expressed little vimentin whereas, the M12 cell line showed a loss of miR-17-3p, suggesting this miRNA might be a tumor suppressor. Restoration of miR-17-3p in the M12 cell line through stable transformation with a miR17-3p

expression plasmid, resulted in a 5-fold increase in miR-17-3p levels, which in turn led to a considerable decrease in vimentin protein expression. In fact, an almost perfect binding of miR-17-3p to vimentin mRNA further confirmed that this mRNA was a strong target for miR-17-3p binding [32].

2.2 Finding additional 3'UTR mRNA targets for miR-17-3p:

In an attempt to find additional mRNA targets for miR-17-3p, three successive passages of M12, M12+miR-17-3p and M12's parental cell line P69 were sent to Dr. Emanuel Petricoin of George Mason University for proteomic analyses via Reverse Phase Protein Microarray (RPMA). The results revealed that the expression of c-Kit, ErbB3, and ErbB2 were the most significantly phosphorylated proteins whose expressions differed between the M12 and P69 cell lines [34]. C-kit was confirmed to be a relevant target for miR 17-3p by western blots, as shown by Seashols, dissertation (2013) [34].

2.3 Mir-17-3p a potential target for 3'UTR ErbB2:

Together with c-KIT, RPMA proteomic analyses revealed that ErbB2 and ErbB3 protein levels were significantly up-regulated, a 1.7- and 2.0- fold increase respectively from the P69 to M12 cell lines. Interestingly, there was no significant change observed in ErbB2 levels from the M12 to M12+miR17-3p cell line, but with a phospho-specific antibody to ErbB2, a 1.6-fold decrease was obtained. It was felt these inconsistent results warranted further investigation to determine if ErbB3 and/or ErbB2 actually contain miR17-3p binding sites or is regulation being exerted indirectly through an unknown kinase which in turn regulates ErbB2 activity. Since luciferase constructs containing ErbB2's 3'UTR were readily available, we started with ErbB2 as it did exhibit a significant change in expression in at least one direction, i.e., p69 to M12 [40].

2.4 Epidermal growth factor receptor (ErbB2):

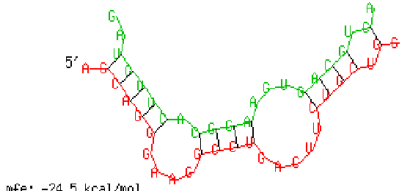
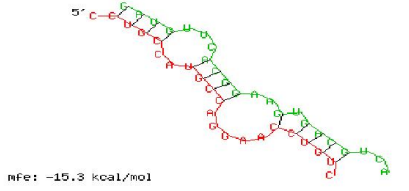
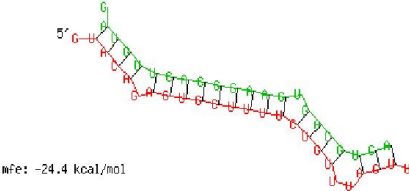
The ErbB2 (also known as HER2, neu, and NGL) belongs to the epidermal growth factor receptor (EGFR) family [41]. Located on chromosome 17q12 from base pairs 39,688,139 to 39,728,661 [40], this gene encodes a 185-kDa transmembrane glycoprotein with phospho-tyrosine activity. Overexpression of this gene has been known to enhance metastasis-related properties such as invasion, angiogenesis, and prolonged growth of cancerous cells, which ultimately leads to metastases. Overexpression of this gene also causes cancerous cells to gain resistance to various chemotherapies and radiation, resulting in a poor response to cancer treatment [41]. If miR-17-3p either by direct or indirect binding can reduce the activity of ErbB2 in cancerous cells, it might represent a relevant therapy for slowing growth, stopping metastasis and eventually enhancing the elimination of cancerous cells through apoptosis.

2.5 Computational analyses of ErbB2 and miR-17-3p:

The miR-17-3p sequence (5'ACUGCAGUGAAGGCACUUGUAG 3') was obtained from miRBase [25] and the sequence of ErbB2's 3'UTR was obtained from the study conducted by Scott *et al* [38]. The sequence of ErbB2's 3'UTR and miR-17-3p was analyzed by RNA-hybrid [37], a structural hybridization prediction tool for measuring the minimum free energy (mfe, or ΔG) based on the alignment of miRNA to mRNA. Surprisingly, three potential miR-17-3p binding sites were observed in ErbB2's 3'UTR of 576 nts in length. The first miR-17-3p site was located at position 44, with a second and third site located at positions 110, and 451 respectively. Table 2-1 depicts the structural interaction of miR-17-3p to the 3' UTR of ErbB2. Based on the proposed strong binding of miR-17-3p to ErbB2's 3'UTR, as suggested by the favorable mfe, it is possible

that one or more of these sites represent favorable binding sites for miR-17-3p which warrants experimental validation.

Table 2-1: The three potential miR-17-3p binding sites (at 5'-position 44, 110, and 451 all extending to 576) within the 3'UTR of ErbB2 from RNA-hybrid software. The value of ΔG indicates the strength of binding between a miRNA and its target mRNA. A low ΔG value indicates a stronger binding affinity. The miR-17-3p binding sites at position 44 and 451 exhibit the strongest binding affinities with a nearly identical ΔG value of -24 kcal/mol.

<p>A: Mir-17-3p site located on position 44</p> <p>MFE: -24.5 kcal/mol</p> <p>Position 44</p> <pre> ErbB2 5' A GAAG GACUU U G 3' GCAGG GCCU CUGC G UGUUC CGGA GACG C miRNA 3' GA A AGU U A 5' </pre>	 <p>MFE: -24.5 kcal/mol</p>
<p>B: Mir-17-3p site located on position 110</p> <p>MFE: -15.3 kcal/mol</p> <p>Position 110</p> <pre> ErbB2 5' C CA AGGAA C 3' CUGC UGCC C CUGU GAUG ACGG G GACG miRNA 3' UUC AA U UCA 5' </pre>	 <p>MFE: -15.3 kcal/mol</p>
<p>C: Mir-17-3p site located on position 451</p> <p>MFE: -24.4 kcal/mol</p> <p>Position 451</p> <pre> ErbB2 5' G G UU U 3' UACA AGUCUUUU CUGU AGU AUGU UCACGGAAG GACG UCA miRNA 3' G U U 5' </pre>	 <p>MFE: -24.4 kcal/mol</p>

2.6 Elucidation of possible miR-17-3p binding sites between 44/576, 110/576, and 451/576 within the 3'UTR of ErbB2:

To determine which site(s) within the 3'UTR of ErbB2 was being repressed the most by miR-17-3p binding, successive 5'-end deletions were prepared as seen in figure 2-1. The construct 44/576 retains all three possible binding sites at positions 44, 110 and 451. Whereas, construct 110/576 lacks site # 1 and finally construct 451/576 retains only site # 3.

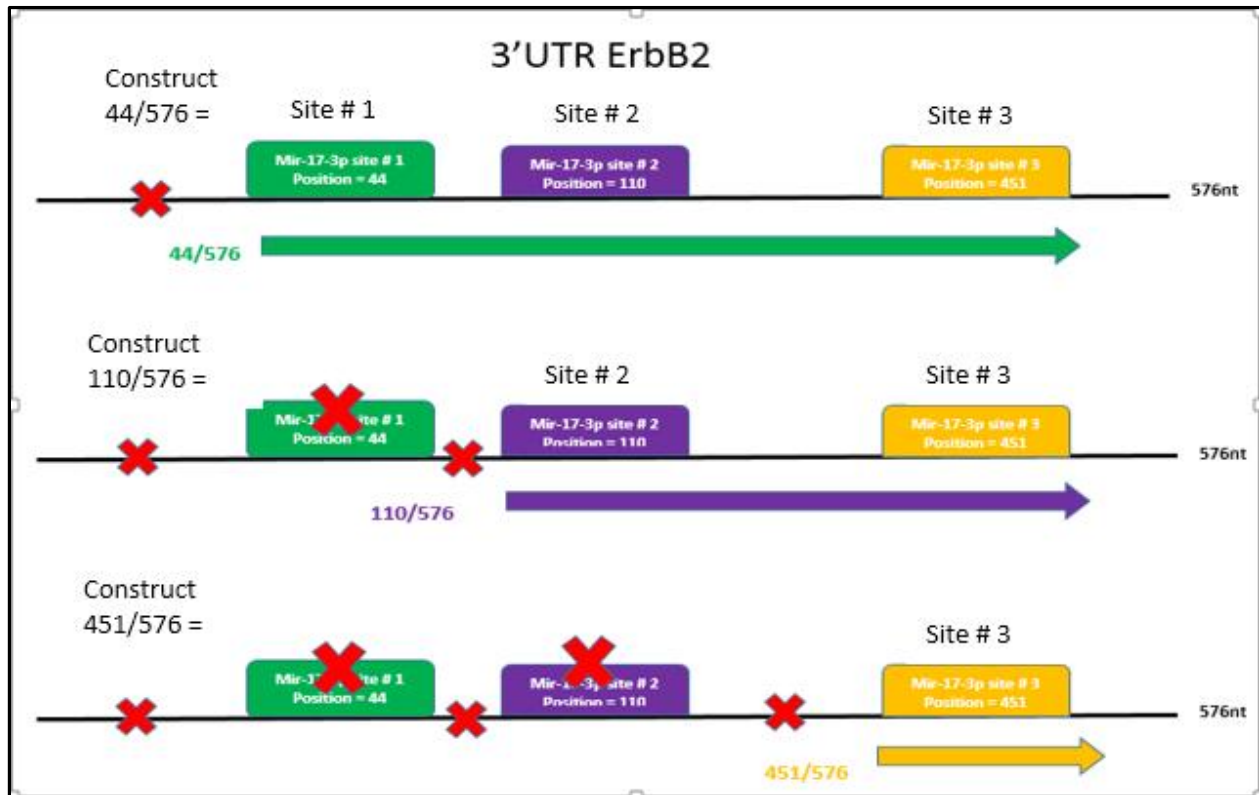


Figure 2-1: The 3'UTR of ErbB2 showing the three binding sites of miR-17-3p at positions 44, 110, and 451 extending to the common 3'-end at nucleotide 576. The arrows are color coordinated with the boxes representing each site and depict how 5'-deletions of each position were prepared. The sequence of the 3'UTR from ErbB2 was from Scot *et al* [38], whereas the miR-17-3p sequences were obtained from miRBase [25]. MiR-17-3p binding sites on the 3'UTR of ErbB2 were deduced using RNA-hybrid software [37]. Each of the miR-17-3p sites were constructed by deleting sequences upstream of each position, but maintaining downstream sequences. Thus, construct 44/576 retained all three possible binding sites, whereas construct 110/576 had deleted site # 1, and finally construct 451/576 retained only site # 3.

2.7 The Goals of this chapter are as follows:

1. Determine which site(s) on the 3'UTR of ErbB2 repressed the most and contributed to the low level of ErbB2 in M12+miR-17-3p cell line.
2. To study the third miR-17-3p site (451/576) on the 3'UTR of ErbB2. This is the only site whose effect can be monitored independently, and not be influenced by other two sites, since they are deleted.

2.8 Methods and Materials

Maintenance of cells

M12 and M12+miRNA media preparation:

In 500 ml of RPMI 1640 supplemented with L-glutamine (Caisson Labs, North Logan, UT), 5% fetal bovine serum (25ml), 250ul of 2000X ITS which consists of 5ug/ml Insulin, 5ug/ml Transferrin, and 5ug/ml Selenium from Collaborative Research, Bedford, MA and 500ul of Gentamycin [0.05 mg/mL] was added as a growth media for the M12 cell line. M12 cells that were stably transformed with the p-SIREN plasmid vector expressing miR-17-3p were maintained using reagents mentioned above and selected with an additional drug, 125ul of puromycin [100ng/ml] [32].

Cell culture:

Cells were cultured in a 75cm² flask and passaged when they reached 65-70% confluency. No cells were cultured above 27 passages. M12 cells remained adherent to the surface of the flask. To pass cells, 2-3mls of pre-heated 0.025% Trypsin-EDTA (Gibco-Life Technologies, Carisbad, CA), were added to the cells and incubated at 37°C for 5-7 minutes. To inactivate trypsin, 8 mls of M12 media (for the M12 cell line) or M12+miRNA media (for M12+miR cell line containing maintenance level of puromycin) were mixed with the trypsin/cell solution and cells pelleted by centrifugation at 1500xg for 5 minutes. The cell pellet was either passaged, or plated based on experimental requirement.

PCR and Cloning of 3' UTR construct:

The 75nt long site# 3 at position 451

(**GAGCTC**TTTGTACAGAGTGCTTTTCTGTTTAGTTTTACTTTTTTTGTTTTGTTTTTTTAAAGATGAAATAAAGACCCAGG**CTAGA**) was amplified via Polymerase chain reaction (PCR). The forward primer was ordered from Invitrogen, Grand Island, NY, USA, and PCR was programmed as follows: 1 cycle at 95°C for 2 min; 30 cycles at 95°C for 30 seconds, followed by 65°C for 30 seconds, 72°C for 1 minute; 1 cycle at 72°C for 10 minutes; stored at 6°C until removed from the machine. The PCR product was directionally cloned into the pmiR-Glo vector (Promega Corporation, Madison, WI) digested with SacI (red) and XbaI (purple) restriction enzymes within the multiple cloning site (MCS) downstream of the luciferase gene using standard digestion and ligation procedures. Upon successful transformation into *Escherichia coli*, and sequence confirmation from Virginia Commonwealth University (VCU) Nucleic Acids Research Facilities (NARF) [58], the plasmid was purified via Qiagen midi-prep kit and used for transfections into the M12+miR-17-3p cell line.

Transfection:

After culturing the cells, the pellet was suspended in 11ml of PBS from which 1mL was used for counting the cells in a Coulter Counter® Analyzer (Beckman Coulter, Inc., Brea, CA) or Vi-Cell™ XR Cell Viability Analyzer (Beckman Coulter, Inc.) using a trypan blue solution. The remaining 10mL of PBS was centrifuged at 1500xg for 5 mins and the pellet re-suspended in the appropriate volume of media to yield 25,000 cells/500ul aliquot of M12 or M12+miR-17-3p cells and plated in a 24-well plate. Each plate was left for incubation at 37°C for 24 hours. Transfections were performed in triplicate. The following day a master mixture was prepared containing per well, pmiR-Glo plasmid (500 ng of vector containing different fragments of ErbB2's 3'UTR as

indicated, 2uL of TransIT®-LT Transfection Reagent (Mirus BIO LLC, Madison, WI, USA), and 50uL of RPMI. The master mix was incubated at room temperature for a minimum of 30 mins. Meanwhile, the media from the previous day was replaced with 500uL of fresh media. The master mix was added to each well drop wise and the plate was incubated at 37°C for 24-48 hours. After incubation the cell media was aspirated and the attached cells washed with 500uL of PBS. Passive lysis buffer (1X) (Promega Corporation) was added to each well (100uL) and allowed to rock back and forth on a platform shaker for 15 min. The cells/buffer solution were scraped off the well and stored in eppendorf tubes at -20°C until the luciferase assay was conducted.

Luciferase Assay:

The luciferase assay was conducted using the Dual-Luciferase Reporter Assay System (Promega Corporation). For this assay the renilla control substrate was prepared, by diluting the Stop & Glo substrate, 1:50 in Stop & Glo solution. One ml of Luciferase Assay Reagent II (LAR II) aliquots was prepared by suspending the luciferine powder in 10 ml luciferase substrate. Once prepared, Stop & Glo (renilla) and LAR II luciferase activities of the cell lysates were measured in the GloMax® 20/20 Luminometer (Promega Contribution) [34]. To an aliquot of thawed cell lysate (20ul), 30 ul of LAR II was added and firefly luciferase activity was measured at a wavelength of 560nm. After 3 seconds, 30ul of renilla substrate was added and renilla activity measured at a wavelength of 480nm. Reporter activity is presented as a ratio of firefly luciferase to renilla luciferase activity. Data was analyzed via the Microsoft® Excel software platform, using the average and standard deviation function with a statistical significance of $p \leq 0.05$.

2.9 RESULTS

Three potential miR-17-3p binding sites were predicted within ErbB2's 3'UTR of 576 nts as follows: site #1 at position 44, site #2 at position 110 and site #3 at position 451. To determine which site(s) was essential in reducing the level of ErbB2, 5'-deletion constructs of these sites were prepared as described above. The activity of each construct was assessed in the M12+miR-17-3p cell line since, this cell line expresses miR-17-3p.

The result of transfection and luciferase assay of each construct, 44/576, 110/576, or 451/576, inserted into the p-miR-Glo vector compared to the empty vector, revealed that construct 44/571 (including all three potential binding sites), decreased luciferase activity the most at 48% (0.52 fold decrease), as expected since it included all three possible binding sites. Construct 110/576 (deleting site #1, but containing site #2 and #3) decreased activity 39% (0.61 fold decrease) whereas construct 451/565 (containing only site #3) exhibited a 15% (0.85 fold) decrease. Investigation of each deleted position indicates that removal of site #1 at position 44 did not result in a large difference in luciferase activity, with only a 9% increase. However, an increase of 24% in luciferase activity was discovered when site #2 at position 110 was deleted (Figure 2-2). This analysis indicates that site #1 at position 44/576 was not an essential site within the 3'UTR of ErbB2 for miR-17-3p binding, while site #2 at position 110/576 and site #3 at position 451/576 were more essential.

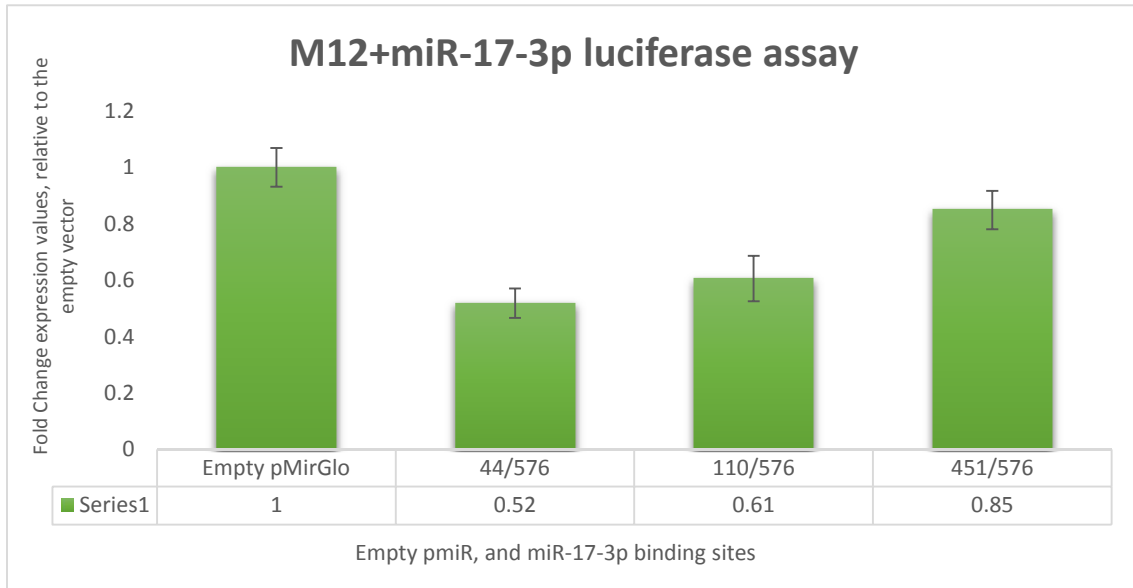


Figure 2-2: Luciferase assays of miR-17-3p’s binding sites within ErbB2’s 3’UTR constructs of 44/576; 110/576; and 451/576 compared to the empty pMiR-Glo vector. Each construct was transfected into the M12+miR-17-3p cell line and firefly luciferase activity measured at wavelength of 560nm versus Stop & Glo (renilla) at a wavelength of 480nm Expression was normalized to the empty pMirGlo vector. Each construct was analyzed in triplicate ($p \leq 0.05$).

2.10 Discussion

Proteomics analyses showed that ErbB2 was significantly up-regulated in the highly metastatic M12 cell line while no significant change was observed in the M12+miR-17-3p cell line. Yet a 1.6X decrease was observed for pErbB2 protein levels in the M12+miR-17-3p compared to the M12. However, since a change in expression for ErbB2 was observed from P69 to M12 cell line and dysregulation of miR-17-3p was observed in the M12 cell line, there was a possibility that miR-17-3p had a binding site within ErbB2's 3'UTR. Here, three potential miR-17-3p binding sites within ErbB2's 3'UTR were analyzed which together were shown to contribute to a major reduction (48%) in luciferase activity. Surprisingly 5'-deletion mapping indicated that site # 2 might be the most effective in reducing luciferase activity as elimination of this region resulted in a 24% restoration of activity. Site #1 on its own only yielded a 9% increase whereas site #3 yielded 15%. There is no doubt that together all three sites are the most efficient in recruiting miR-17-3p and in the cell would be the most responsive to miRNA regulation of ErbB2 synthesis.

These analyses revealed that site # 2 and somewhat site # 3 are probably the most effective binding sites with ErbB2's 3'UTR for miR-17-3p binding. This conclusion did not confer with the results obtained from RNA-hybrid [37] for deducing the best binding sites for miR-17-3p on ErbB2's 3'UTR based on mfe as a valid measurement of binding efficiency. RNA-hybrid [37] showed site # 1 and site # 3 to be the best binding sites, whereas, luciferase assay conducted in our highly metastatic prostate cancer M12+miR-17-3p cell line, revealed that site # 2 and site # 3 were the best binding sites within the 3'UTR of ErbB2. The variation among these two type of analyses could be due to a variety of reasons. First, in deleting site # 2 a large section of ErbB2's 3'UTR was deleted (from bases 110 to 451) which might contain additional sequences (perhaps a partial

match to miR-17-3p) which independently presents an additional landing platform to assist miR-17-3p binding to site #2. Secondly, there may be additional unknown miRNA binding sites within the 3'UTR of ErbB2, which in our cell line might be indirectly affecting miR-17-3p binding. Our miR screen analyses suggests that many miRNAs were differentially expressed in the various cell lines of our cancer progression model [40]. Therefore, there is a strong possibility that apart from miR-17-3p there are other potential miRNAs that are competing for binding to the 3'UTR of ErbB2. In fact, there may be some undiscovered miRNAs with partial complementarity to site #1 which might be competing with miR-17-3p for binding; thus, contributing to the loss of repression to a site which due to a favorable mfe should be a strong binder for miR17-3p.

To further understand the perceived strong binding by site #2, additional experimentation is required. Since 5'-end deletion mapping generates a different 5'-end for each construct relevant to a fixed 3'-end, the results can sometimes be misleading. A better analysis would be to keep the same 3'-UTR sequence and mutate each potential site by site directed mutagenesis within that fixed 3'UTR segment. In this manner those specific bases complimentary to the seed region of miR-17-3p would be individually mutated for each site and then in various combinations, i.e., site #1 + site #2 or site #2 + site #3 and so on, to determine exactly which site on its own is the most effective in binding miR-17-3p.

In conclusion, these analyses confirmed that the 3'UTR of ErbB2 contains active sites for miR-17-3p binding. This result confirms our proteomic analysis suggesting that ErbB2 is a direct target for miR-17-3p binding. Supporting the increase in ErbB2 protein levels in P69 to M12 cell line. The loss of miR-17-3p as a tumor suppressor in relevant prostate cancer cell lines and in tumor tissue by laser capture microdissection, suggests how the expression of ErbB2 is enhanced. ErbB2 is an important molecule in cancer. This gene belongs to the EGF family, a growth factor

receptor protein that is located on the cell surface enabling the activated receptor complex to relay signals in the cell. These signaling transduction pathways activate certain genes that promote cell growth [39]. Overexpression of ErbB2 has been linked to the metastatic potential of cancer cells and eventually these cancerous cells gain resistance to treatments such as radiation and chemotherapy. The network model analyses, previously conducted in our laboratory, revealed that ErbB2 proteins have highly connected nodes in prostate cancer, suggesting that these proteins are key regulators of development and progression of prostate cancer [40].

Chapter 3:
**The significance of hsa-mir-299-5p, and possibly hsa-mir-299-3p,
contributing to prostate cancer progression**

3.1 Introduction and previous work:

Hsa-miR-299-5p and hsa-miR-299-3p are located on chromosome 14 (14q 32.31) from 101490131 base pairs to 101490193 base pairs [44]. Both these miRNAs differed in expression in the human panel-miR analyses and single-miR analyses. Hsa-miR-299-5p was found to be a tumor suppressor exhibiting a 0.02-fold reduction in the panel analyses whereas a value for the single-miR analyses, at that time, was not obtained. An overexpression of 11.4-fold, was observed for hsa-miR-299-3p in the panel analyses, whereas a conflicting 0.035-fold loss of expression was observed in the single-miR analyses (Figure 3-i) [34]. Due to vast differences observed between panel analyses and single miR analyses, it felt further investigation on these miRNAs was important.

Expression differences between panel and single-miR analysis								
	133a	133b	9	622	199a-3p	147b	147	
Panel	37.9	38.9	6.7	147.7	6.43	201.4	27.9	
Single-miR	6.7	2.5	43.3	0.374	1.514	16.9	0.138	
	146a	488	221	127-3p	127-5p	299-3p	299-5p	144
Panel	0.459	0.526	0.783	0.002	0.344	11.4	0.02	0.374
Single-miR	0.021	0.039	1.39	0.003	0.164	0.035	---	---

Figure 3-i: Differential expressions observed between miR-299-3p and -5p for panel and single-miR analyses. In these analyses data was normalized against the global mean for panel and against RNU48 for single-miR analyses. M12 expression level was measured relative to the expression level of its parental P69 cell line for single miR analysis. MiR-299-3p was overexpressed in the panel analyses; however, highly reduced expression was observed for the single-miR analyses. Mir-299-5p appeared to be down-regulated in panel analyses, whereas no expression value was measured for single-miR analysis.

[Figure adapted from Seashols, dissertation (2013) [34]]

RNA was extracted via laser capture microdissection (LCM) from 5 patient biopsies, which were obtained by radical prostatectomies at different stages of prostate cancer. MiR Locked Nucleic Acid (LNA) RT-qPCR analyses of these patient samples showed that despite vast variability among patient samples some miRNAs remained consistent in their nature, while others varied. As mentioned in chapter 1, miR-9 and miR-147b, in particular, remained strongly consistent as oncomiRs in the model cell lines as well as in patient tumor samples. MiR-299-3p and miR-199-3p, remained consistently low as tumor suppressors in model cell lines as well as patient tumor samples [Figure 3ii-B and C] [34]. However, MiR-299-5p is an unusual (or maybe irregular) miRNA as a number of fluctuations in its expression were observed, first during single-miR analyses and then during patient samples. Two individual experiments were performed on patient samples to measure the expression level of this miRNA. MiR-299-5p was found to be up-regulated in the first experiment consisting of three patient samples where miR-299 was detected (Figure 3ii-A). However, in the other subsequent experiments miR-299-5p was down-regulated (Figure 3ii, B and C). In the literature miR-299-5p has been reported to be down regulated in progressing from normal to localized tumor to metastatic cases of prostate cancer [44]. The great variation in the expression of miR-299-5p versus the consistent expression of miR-299-3p both deriving from the same miR-299 precursor, demands further study.

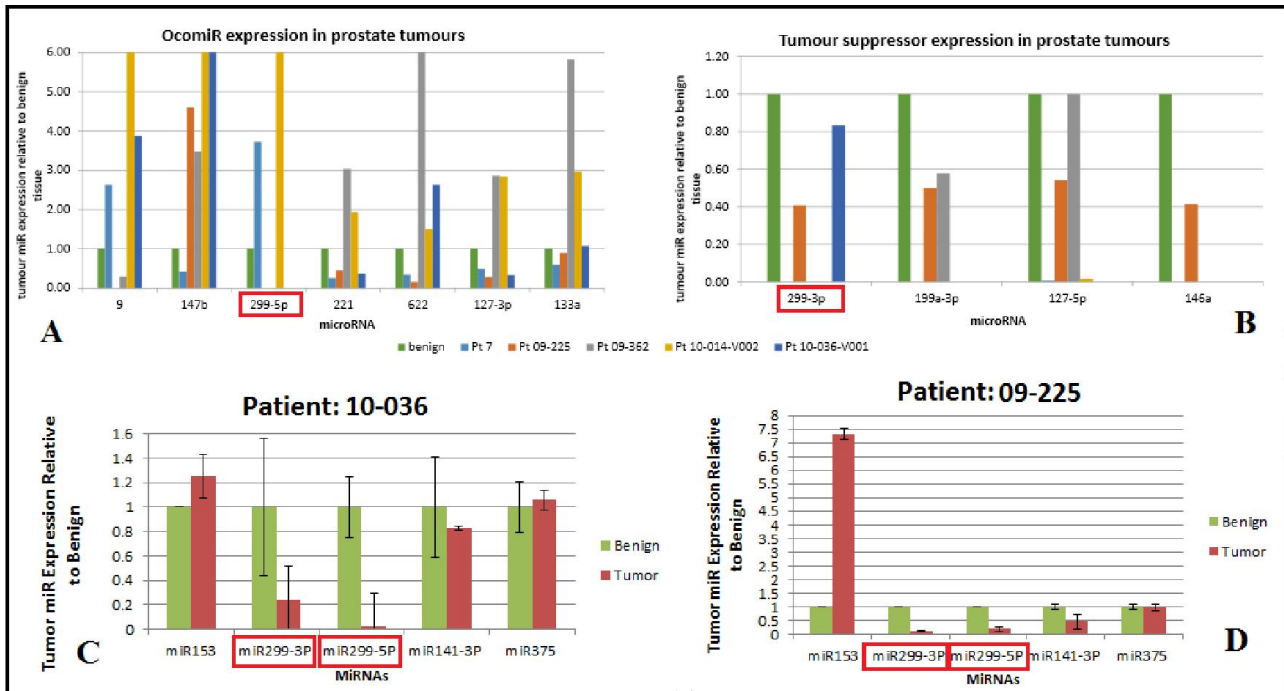


Figure 3-ii: The expression levels of miR-299-3p and -5p observed for prostate cancer patients in benign and tumor samples. Two individual experiments were performed on patient samples. A-B: First experiment conducted on patient samples. The expression levels relative to benign revealed that miR-299-5p was up-regulated in two patient samples, classifying it as an oncomiR. MiR-299-3p was down-regulated in two patient samples, classifying it as a tumor suppressor. C-D: Second repeats conducted on patient samples. The expression levels relative to benign, revealed that while miR-299-3p was consistently down-regulated, so was miR-299-5p, now classifying it as a tumor suppressor. Another discrepancy seen here is for patient 10-036, in part A, miR-299-3p is seen as an oncomiR and in part C, miR-299-5p is seen as a tumor suppressor. Whereas, for patient 09-225, in part A, miR-299-5p is not detected, and in part D, miR-299-5p is seen as a tumor suppressor.

[Figure 3-ii-A-B were adapted from Seashols, dissertation [34], figure 3-ii C-D were adapted from former laboratory student J.H.]

3.2 Hsa-miR-299-5p

Over the last years miR-299-5p has gained importance in a variety of developmental and disease related studies. For example, one report found a link between miR-299-5p and pregnancy. Together with other miRNAs, miR-299-5p was detected at a high level before delivery with a drop to null after delivery [45]. Another report studied the regulation and role of up-regulated miRNAs localized to the chromosome 14q32 miRNA cluster, in interstitial lung disease and idiopathic pulmonary fibrosis (IPF), found that both miR-299-5p and miR-299-3p among other miRNAs were significantly up-regulated. [46].

For prostate, targets regulated by miR-299 are not well known; thus, it is important to find relevant targets which correlate with prostate cancer, and determine if changes in miR-299-5p or -3p expression affect the presence of proposed target mRNA/protein levels.

3.3 Osteopontin (OPN):

The literature suggests that osteopontin (OPN) is targeted by miR-299-5p. OPN is known as secreted phosphoprotein 1 (SPP1), a secreted glycosylated phosphoprotein that consists of 2% non-collagenous proteins in bone. It is present in all body fluids and an increase in its expression is known to correlate with tumor invasion and metastasis in prostate, colon, lung, breast and other cancers [47]. In a study that was performed to determine if OPN evokes a humoral immune response in PCa patients and if anti OPN antibodies can be used to differentiate PCa patients from healthy individuals, it was found that OPN was over expressed in prostate cancer and contributed to the progression of the disease. Healthy donors, biopsy proven PCa patients, and patients with benign prostate hyperplasia (BPH) were tested using recombinant human OPN. The frequency of anti-OPN antibodies found in the PCa samples was higher than in BPH and healthy individuals.

Based on these results the study proposed that anti-OPN antibodies may be used as an early serum marker for PCa [48]. Similarly, the importance of OPN was determined in the MCF10DCIS cell line for breast cancer. MCF10 is a model cell line consisting of normal immortalized breast epithelial cells [49]. The MCF10DCIS cell line was established from a xenograft lesion originating from premalignant MCF10AT cells injected into combined immune-deficient mice. This study found that when MCF10DCIS cells were grown as spheroids the oncogenic protein OPN was secreted in high levels thereby contributing to tumorigenicity. It was shown that OPN expression was down-regulated in MCF10DCIS spheroids due to being targeted by hsa-miR-299-5p. It is possible that the 3'-UTR of OPN may contain a binding site that is recognized by hsa-miR-299-5p [50].

3.4 Binding of miR-299-5p to the 3'UTR of OPN:

There is not much literature that provides evidence of the 3'UTR of OPN being a proven target for hsa-miR-299-5p. However, one study conducted experiments on the multicellular spheroid-forming sub-population (SFC) of the following human breast cancer cells, MCF-7, MCF10AT, and MCF10DCIS compared to COS-7, to determine if the 3'UTR of OPN contains a binding site for hsa-miR-299-5p [50]. Among various experiments conducted, one experiment determined that OPN indeed evoked tumorigenicity in breast cancer cells. The results showed that upon abrogation of OPN expression by adding a neutralizing anti-OPN antibody, the proliferation rate of the MCF7-SFC and MCF10DCIS-SFC cells was reduced to 40% and 60%, respectively [50]. In order to confirm whether the 3'UTR of OPN was bound by hsa-miR-299-5p, the MCF10AT cells and MCF10DCIS cells were transfected with an anti-miRNA-inhibitor specific for hsa-miR-299-5p. Initially, upon transfection with 40 nM of anti-miRNA inhibitor, the levels of hsa-miR-299-5p remained unaltered; however post 48 hours of treatment, the levels of OPN up-

regulated and increased even further in correlation with increasing the amount of inhibitor to 60nM and ultimately 100nM [50]. This result confirms that with increasing concentrations of anti-miRNA inhibitor, the ability of hsa-miR-299-5p to bind to OPN's 3'UTR decreased resulting in a concomitant increase in OPN protein levels. Lastly, the ability of hsa-miR-299-5p to bind to the 3'UTR of OPN was tested in two cell systems, COS-7 and MCF10A. The 3'UTR of OPN mRNA was cloned downstream of a luciferase reporter in the pmiR-Report vector to generate pmiR-Report-OPN and co-transfected with hsa-miR-299-5p into COS-7 or MCF10A cells resulting in a 42% and 30% reduction in luciferase activity, respectively [51]. This experiment confirmed that the 3'UTR of OPN contains a binding site for hsa-miR-299-5p.

3.5 Hsa-miR-299-3p

Studies have shown that the androgen receptor (AR) is targeted by hsa-miR-299-3p [52]. Initially, androgen is important for the growth of prostate cancer and one way to control PCa is to inhibit AR; however, overtime prostate cancer becomes AR independent and continues to thrive. AR is expressed in all stages of prostate cancer hence miRNAs could be used to control the development of PCa. Ostling *et al.* found that the AR has an unusually long 3'UTR (>6Kb) which is regulated by 13 miRNAs. Among those miRNAs, transfection of a miR-299-3p mimic resulted in a decrease of AR expression as measured with the relevant luciferase:AR-3'-UTR fusion construct [52]. Since AR is critical for progression of prostate cancer, an initial treatment for prostate cancer is removal of AR. In normal prostate, AR promotes survival and differentiation, but during prostate cancer development it becomes an inducer of uncontrolled cell growth. With hsa-miR-299-3p being a potential tumor suppressor, its restored expression in prostate cancer and in relevant cell lines *in vitro* could possibly result in decrease expression of AR, therefore inhibiting uncontrolled prostate cancer cell growth.

3.6 Experimental Aims

To the best of our knowledge, no study has published the effect of hsa-miR-299-5p binding to the 3'-UTR of OPN in prostate cancer. Furthermore, no study has engineered a prostate cancer cell line to stably overexpress miR-299-5p or miR-299-3p and monitored the consequences to cell behavior. Previous studies have only co-transfected the miRNA with its known luciferase reporter construct fused to the target 3'UTR (luc:3'UTR) to validate a decrease in luciferase activity dependent on increased miRNA content. Henceforth, the goals of this research were to:

1. Restore the expression of miR-299 in the M12 cell line, using the precursor hsa-miR-299 expression plasmid from GeneCopoeia [53] to create a stable transformant to be named M12+miR-299.
2. Conduct q-RT-PCR on the M12 cell line and M12+miR-299 cell line to determine which mature hsa-miR-299 (-5p or -3p or both) was being produced in our prostate cancer cell line.
3. Confirm the functional consequences of altered miR-299 expression on target mRNA expressions using the pmiR-Glo luciferase: fusion plasmid construct (Luc:OPN3'UTR)
4. Determine if repression of luciferase activity is observed dependent on the inclusion of OPN's 3'UTR then ultimately construct mutants of relevant nucleotides within the OPN's 3'UTR target and re-assay.
5. Conduct migration, invasion and proliferation assays on M12 vs. M12+miR-299 to determine the effect of altered miR-299 expression on migratory abilities and cell behavior.

3.7 Methods and Materials:

M12 and M12+miRNA media preparation:

In 500 ml of RPMI 1640 supplemented with L-glutamine (Caisson Labs, North Logan, UT), 5% fetal bovine serum (25ml), 250ul of 2000X ITS which consists of 5ug/ml Insulin, 5ug/ml Transferrin, and 5ug/ml Selenium from Collaborative Research, Bedford, MA and 500ul of Gentamycin [0.05 mg/mL] was added as a growth media for the M12 cell line. M12 cells that were stably transformed with the p-miRGlo plasmid vector expressing miR-299-5p were maintained using reagents mentioned above and selected with an additional drug, 125ul of puromycin [100ng/ml] [32].

Preparing to insert the miR-299 precursor in the M12 cell line:

The hsa-miR-299 precursor gene, as a stem-loop expression plasmid was obtained from (GeneCopoeia™,Rockville,MD).

(5'AAGAAAUGGUUUACCGUCCACAUACAUUUUGAAUAUGUAUGUGGGAUGGUA AACCGCUUCUU3') as a bacterial broth. *Escherichia coli* was transformed with 25ul of the bacterial broth, plated on LB agar plates containing the antibiotic ampicillin, colonies picked, and plasmid purified by standard mini-prep purification methods. After confirming the sequence of the precursor through Virginia Commonwealth University, Nucleic Acid Research Foundation (VCU-NARF) [58] a larger scale, Midiprep (Qiagen) was performed to extract DNA suitable for transfecting the miR-299 plasmid into the M12 prostate cancer cell line.

Stable Transfection

In 25 cm² flask 1.5 x 10⁵ M12 cells were plated in the appropriate serum containing media for 24 hours. The following day a mixture of plasmid DNA (6.5 µg) containing the miR-299 gene insert, 19.5 µl of TransIT®-LT Transfection Reagent (Mirus BIO LLC, Madison, WI, USA) and 650 µl of serum/additive free RPMI media was mixed together and left for incubation at room temperature for 30 minutes. The solution was then added drop-wise to each flask, gently rocked back and forth, and placed in the 37°C incubator for 42 hours. The old serum and additives free RPMI media was aspirated off and stable transformants selected using cell culture media containing 200 ng/mL of Puromycin for four weeks. After four weeks, the cells, now known as M12+miR-299, were maintained with cell culture media containing 100ng/mL puromycin (maintenance level).

Maintenance of cells

Cell culture:

Cells were cultured in a 75cm² flask and passaged when they reached 65-70% confluency. No cells were cultured above 27 passages. M12 cells remained adherent to the surface of the flask. To pass cells, 2-3mls of pre-heated 0.025% Trypsin-EDTA (Gibco-Life Technologies, Carisbad, CA), were added to the cells and incubated at 37°C for 5-7 minutes. To inactivate trypsin, 8 mls of M12 media (for the M12 cell line) or M12+miRNA media (for M12+miR cell line containing maintenance level of puromycin) were mixed with the trypsin/cell solution and cells pelleted by centrifugation at 1500xg for 5 minutes. The cell pellet was either passaged, or plated based on experimental requirement. *Refer below to “cell pellet preparation” in order to extract RNA.*

Cell pellet preparation:

Cell pellets (either M12 or M12+miR-299) were washed with 10 mL of PBS and re-centrifuged at 8000xg for 5 minutes and supernatant removed and the cell pellet washed one more time with PBS as above. The cell pellets were flash frozen in liquid nitrogen and stored at -80°C for at least 24 hours.

RNA extraction from cell pellets:

RNA was extracted using the miRVana™ RNA isolation kit and recommended protocol (Ambion-Life Technologies, Carlsbad, CA). RNA was eluted in 50 µl of Elution Buffer and stored in -80°C. RNA concentration was measured using the NanoDrop ND-2000 Spectrophotometer (Thermo-Fisher Scientific, Inc., Waltham, MA) [34].

cDNA synthesis:

For cDNA synthesis and mRNA quantification the QScript™ cDNA synthesis kit (Quanta Biosciences, Gaithersburg, MD) was used. cDNA was prepared in thin-walled PCR tubes using 1-7µl of RNA [1 µg down to 10 pg], 2µl of Poly(A) Tailing Buffer (5X), and 1µl of Poly(A) Polymerase. The mixtures were vortexed and centrifuged for 10sec to collect the contents. The tubes were placed in the thermal cycler programmed as follows: 1 cycle at 25°C, 1 min; 1 cycle at 37°C, 40 min; 1 cycle at 70°C, 5min, followed by 6°C, 5-10min; During the hold period 10µl of Poly(A) Tailing reaction, 9µl of microRNA cDNA Reaction Mix and 1µl of qScript Reverse Transcriptase was added to each tube. The tubes were placed back in the thermal cycler program resuming from 6°C and moving to 1 cycle at 42°C for 20 minutes and 85°C for 5 min. The tubes were stored at -80°C for further analyses.

QPCR:

In a 96-well-plate miRNA content of cDNA from M12 and M12+miR-299 was analyzed. Each experiment was performed in triplicate. Five master mixes were prepared containing: Quanta Biosciences, Gaithersburg, MD, 10uM (0.25uL) - 299-3p primer, 299-5p primer, RNU-48, miR-375 and miR-141, 6.25 uL SyberGreen, 2uL cDNA, and 4uL H₂O. The QPCR instrument was programmed as Step 1: 1 cycle at 95°C for 2-3 minutes; step 2: 40 cycles at 95°C for 15seconds, 60°C for 15 seconds, and 70°C for 30 seconds; step 3, 1 cycle at 95°C for 15 seconds, 60°C for 30 seconds, and 95°C for 15seconds.

DNA Synthesis and cloning

For the miR-299 target, osteopontin (OPN), two strands from the 3'UTR of OPN were synthesized as follows (Invitrogen, Grand Island, NY, USA).

The 3'UTR OPN was found in the National Center for Biotechnology Information (NCBI) accession number: NM_001040058.1.

The 1st strand OPN 9 (insert = 56nt) and 2nd strand OPN 13 (insert = 55nt) being

1) 5' **CGCGGCCGC**CCTTACAACAAATACCCAGATGCTGTGGCCACATGGCTAAACCCTGACCCATCTCAGT3'

2) 5' CAGTTTGTGGCTTCATGGAAACTCCCTGTAAACTAAAAGCTTCAGGGTTATGTCTAG**CGCGGCCGCT** 3'

The red, blue and green are SACI, NOTI and ACCI restriction enzymes, respectively used for directional cloning digested into the pMir-Glo vector. Each oligonucleotide was re-suspended at 100 uM. For annealing 9 ul of the forward and reverse strands was mixed with 2ul of 10X annealing buffer and heated at 95°C for 20 minutes. The mixture was allowed to cool at room temperature for 120 minutes, and based on experimental requirements was either stored at -20C or immediately used for ligation. OPN inserts were ligated into the p-Mir-Glo vector cut with SACI and ACCI restriction enzymes. The ligates were allowed to sit at room temperature for minimum

of 30 minutes and based on further experiments, was either used for transformation or stored at -20C. The standard cloning procedure was carried out and 5ul of the ligated products was transformed into DH5 α *Escherichia coli* cells and 150ul or 250ul of the transformants were plated on agar plates and selected with the drug ampicillin. The standard miniprep following the manufacturer's protocol (Qiagen) was carried out and DNA was sent to VCU-NARF [58] for confirmatory sequencing. Upon confirmation, midiprep DNA (Qiagen) based on the manufacturer's protocol was performed to extract a plasmid suitable for transfection. For transfections only midiprep plasmid products were used due to the high purity and concentration of DNA obtained.

Transfection of 3'UTR of OPN into M12 and M12+miR-299 cell lines

After culturing the cells, the pellet was suspended in 11ml of PBS from which 1mL was used for counting the cells in Coulter Counter® Analyzer (Beckman Coulter, inc, Brea, CA) or Vi-Cell™ XR Cell Viability Analyzer (Beckman Coulter, Inc) [34] using trypan blue solution. The remaining 10mL of PBS was centrifuged at 1500xg for 5 mins and pellets re-suspended in the appropriate volume of media to yield 25,000 cells/500ul. Aliquots of M12 and M12+miR-299 cells were plated in a 24 well plate and left for incubation at 37°C for 24 hours. Each sample was performed in triplicate. The following day a master mixture was prepared, containing per well pmiR-Glo plasmid (500 ng of vector containing the 3'UTR of OPN), 2uL of TransIT®-LT Transfection Reagent (Mirus BIO LLC, Madison, WI, USA), and 50uL of RPMI. The master mix was incubated at room temperature for a minimum of 30 min. Meanwhile, the media from the previous day was replaced with fresh media. The master mix was added to each well drop wise and the plate incubated at 37°C for 24-48 hours. After incubation, the cell media was aspirated and the attached cells washed with 500ul of PBS. 100ul of 1X Passive Lysis Buffer (Promega Corporation) was

added to each well and allowed to rock back and forth on a platform shaker for 15 min. The cells/buffer solution was scraped off the well and stored in eppendorf tubes at -20°C until needed to conduct luciferase assay.

Luciferase Assay:

The luciferase assay was conducted using the Dual-Luciferase Reporter Assay System (Promega Corporation). For this assay the renilla control substrate was prepared, by diluting the Stop & Glo substrate, 1:50 in Stop & Glo solution. One ml of Luciferase Assay Reagent II (LAR II) aliquots was prepared by suspending the luciferine powder in 10 ml luciferase substrate. Once prepared, Stop & Glo (renilla) and LAR II luciferase activities of the cell lysates were measured in the GloMax® 20/20 Luminometer (Promega Contribution) [34]. To an aliquot of thawed cell lysate (20ul), 30ul of LAR II was added and firefly luciferase activity was measured at a wavelength of 560nm. After 3 seconds, 30ul of renilla substrate was added and renilla activity measured at a wavelength of 480nm. Reporter activity is presented as a ratio of firefly luciferase to renilla luciferase activity. Data was analyzed via the Microsoft® Excel software platform, using the average and standard deviation function with a statistical significance of $p \leq 0.05$.

Preparing the mutants of 3'UTR of OPN:

For this analysis, it was felt that the length of the original OPN 3'UTR , needed to be shortened to minimize any possible shift in miR-299-5p binding to downstream sites containing a partial match. Thus, two new sets of original 3'UTR of OPN 9 and OPN 13 created were:

- 1) (OPN 9 insert = 25nt) 5' **CGCGGCCGCTGCTGTGGCCACATGGCTAAACCC**TGT 3'
- 2) (OPN 13 insert = 26nt) 5' **CAGTTTGTGGCTTCATGGAAACTCCCTCGCGGCCGCGT** 3'

The mutants (underlined) created for 1 and 2 are as follows:

- 1) OPN 9: 5' **CGCGGCCGCTG**CTTTCACAACATTGCATTTTTCTGT 3'
- 2) OPN 13: 5'CAGTTAGAGTCTTCATTTTATTTCCCTGCGGCCGCGT 3'

As described above, the same methods for DNA synthesis, cloning, transfections, and luciferase assay were also performed.

Migration and Invasion Assay:

Migration assay was carried out with M12 and M12+miR-299 cell lines. After culturing, cell pellets were suspended in 10ml of PBS from which 1mL was used for counting the cells in Coulter Counter® Analyzer (Beckman Coulter, inc, Brea, CA) or Vi-Cell™ XR Cell Viability Analyzer (Beckman Coulter, Inc) using trypan blue solution. Cells were collected by centrifuged at 1500xg for 5 minutes as described above and re-suspended in RPMI media (no additives added) media in an appropriate volume. For the migration assay 50,000 cells per 200 uL were added to a ThinCert™ TC membrane support insert (Greiner Bio-one BVBA/SPRL, Belgium) and placed in a well in 24-well plate. For the invasion assay before adding the cells, the ThinCert™ membrane support insert was layered with 30ul of a 10% solution of Cultrex® Basement Membrane Extract (Growth Factor Reduced), (R&D Systems®, Minneapolis, MN) mixed in RPMI 1640 medium and left to incubate at 37°C for 1 hour. After 1 hour 125,000 cells per 500ul were added to the membrane support insert. For both assays, as a chemoattractant- 1mL of, above mentioned, M12-media containing 10ng/mL EFG was pipetted into the lower chamber of the well and the plate incubated at 37°C for 20-24 hours. Media from both the membrane and lower chamber was aspirated off and cells fixed by adding 300uL of 0.025% glutaraldehyde in PBS to the membrane and the lower chamber for 20 minutes. The glutaraldehyde was aspirated off and replaced with 0.1% Crystal Violet in 10% EtOH and PBS for a minimum of 30 minutes. The membranes were excised and mounted on a

microscope slide using Permount® (Fisher Scientific, Waltham, MA). Cells were counted on 5 random fields for each replicate, averaged and expressed as relative to the control (M12 cell line).

Proliferation Assay

After culturing the cells, the pellet was suspended in 10ml of PBS from which 1mL was used for counting the cells in Coulter Counter® Analyzer (Beckman Coulter, inc, Brea, CA) or Vi-Cell™ XR Cell Viability Analyzer (Beckman Coulter, Inc) using trypan blue solution. The rest 9mL of PBS was centrifuged at 1500xg for 5 minutes and suspended pellets in raw RPMI (no additives added) media and appropriate volume. Cells were harvested, pelleted and counted as described previously and re-suspended for this assay as 10,000 cells/100 uL for plating in a 96 well plate. Media plus or minus puromycin, but without cells was added to several additional wells to act as a negative control for background fluorescence or M12 or M12+miR-299 cells, were plated in 9 wells and incubated at 37°C for 24, 48, and 72 hours. At each time interval the media was aspirated and replaced with 100uL of fresh cell culture media. 10uL of 12mM MTT stock solution was added to each well, including the negative controls. The unit was incubated at 37°C for 4 hours. The MTT stock was prepared by adding 1mL of PBS to a vial containing 5mg of MTT. The mixture was vortexed or sonicated until dissolved. The resulting MTT solution can be stored for four weeks at 4°C, protected from light. After 4 hours, 85uL of the media was removed from each well and 50uL of DMSO was added and mixed thoroughly. The unit was incubated at 37°C for 10 minutes, and the absorbance read at 540 nm.

3.8 RESULTS:

The RT-qPCR results for M12 and M12+miR-299 (Figure 3-1) showed that miR-299-5p expression increased 18-fold in the M12+miR-299 stable transformants. There was no significant increase in miR-299-3p expression in the M12+miR-299 cells compared to its parental M12 cell line. From this result it can be concluded that the M12+miR-299 stable transformant is exhibiting an altered expression of only miR-299-5p. Also, the expression of other miRs such as miR-375 or miR-141, remained un-altered and exhibited no change in either the M12 or M12+miR-299 cell lines. This confirmed that inadvertently, the expression of other representative miRNAs was NOT being altered by the stable over expression of miR-299-5p.

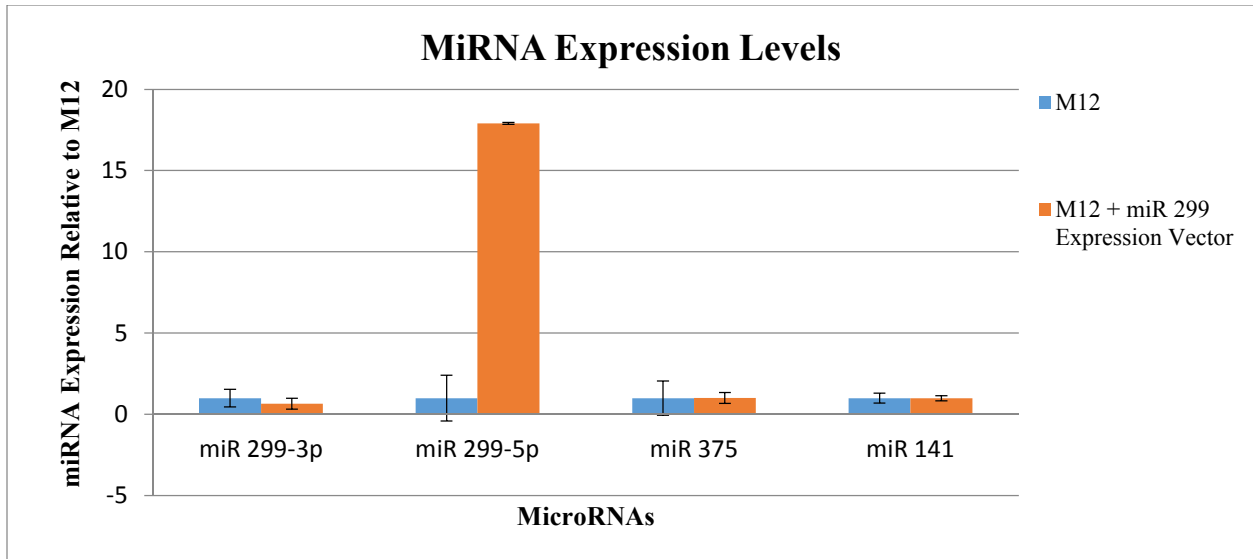
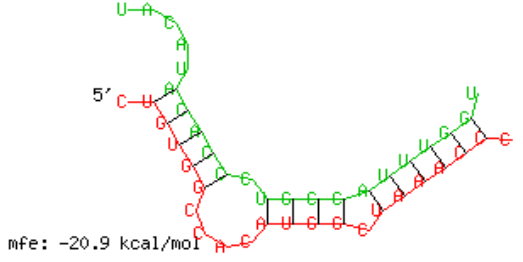
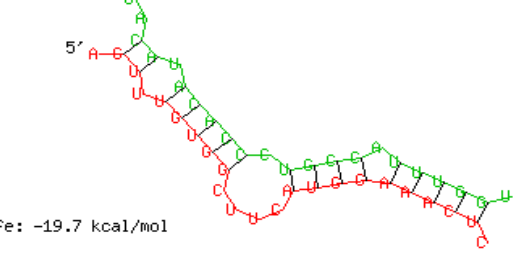


Figure 3-1: The expression levels of miR-299-3p, miR-299-5p, miR-375, and miR-141 in the M12 and M12+miR-299 cell lines. RT-qPCR data was normalized to RNU48. The analyses confirmed that stable transformation of the M12 cell line with a miR-299 mini gene (M12+miR-299-5p) resulted in enhanced expression of only miR-299-5p with no increase in expression of miR-299-3p or no effect on the expression of other unrelated miRNAs. Each sample was analyzed in triplicate ($p \leq 0.05$).

A search of the literature revealed a binding site for miR-299-5p at position 1269-1290 within the 3'UTR of OPN [51]. However, to determine if a better or additional binding sites is present, the 3'UTR OPN was analyzed in RNA-hybrid [37] software. Table 3-1, shows two sequences within OPN's 3'UTR with the best possible matches for miR-299-5p. Among these two, one match (3'UTR OPN 13) coincided to that cited in the literature [51]. The mfe value, for 3'UTR OPN 9 (position 280-301) was -20.9 kcal/mol whereas that for 3'UTR OPN 13 (position 1269-1290) was -19.7 kcal/mol suggesting that 3'UTR OPN 9 might be a better or at least an additional site for miR-299-5p binding. To determine which or if both sites are functional *in vitro*, each 3'UTR site was synthesized, cloned into the pmiR-Glo vector and the resulting luciferase activity analyzed in M12 or M12+miR-299-5p stably transformed cell lines.

Table 3-1: The miR-299 binding on the 3'UTR OPN 9 (56nt) and 3'UTR OPN 13 (55nt). On the left column are miR-299-5p (green sequence), OPN 9 (red sequence) with mfe of -20.9 kcal/mol and OPN 13 (blue sequence) with mfe of -19.7 kcal/mol. The right column shows the binding of miR-299-5p (green sequence) to OPN 9 or OPN 13 (both red sequences).

<p>MFE: -20.9 kcal/mol (OPN 9) <u>Position 281</u> Target 5' C CCAC C C 3' UGUGG AUGG UAAACC ACACC UGCC AUUUGG MIRNA 3' UACAU C U 5'</p>	 <p>mfe: -20.9 kcal/mol</p>
<p>MFE: -19.7 kcal/mol (OPN 13) <u>Position 1269</u> Target 5' A U CUUC C 3' GU UGUGG AUGG AAACU CA ACACC UGCC UUUGG miRNA 3' UA U C A U 5'</p>	 <p>mfe: -19.7 kcal/mol</p>

As anticipated the transient transfection of the luc:3'UTR-OPN9 and luc:3'UTR-OPN 13 plasmids in M12 and M12+miR-299 cell lines exhibited repression for both luc:3'UTR OPNs (figures 3-2A, 3-2B, respectively). The results showed that, in comparison to the empty pmiR- Glo vector in the M12 cell line, OPN 9 binding site repressed luciferase activity 24% whereas OPN 13 repressed only 5%. Concurrently, the M12+miR-299 cell line compared to the empty pmiR-Glo yielded OPN 9 and OPN 13 repression of 31%. The latter result was slightly better (1%) than what was found in the literature for similar transfections in the MCF10AT cell line (30%). Additionally, the repression rates of OPNs measured in M12 vs. M12+miR-299 showed that OPN 9 repressed 7% and OPN 13 repressed 26% more in the M12+miR-299 cell line. Thus, at this point it appeared that OPN 13 containing the miR-299-5p binding site with the slightly better mfe binding credentials exhibited greater inhibition in the M12 cell line. However, in M12 cells engineered to overexpress miR-299-5p (M12+miR-299-5p) both OPN sites were comparable in capturing miR-299-5p binding.

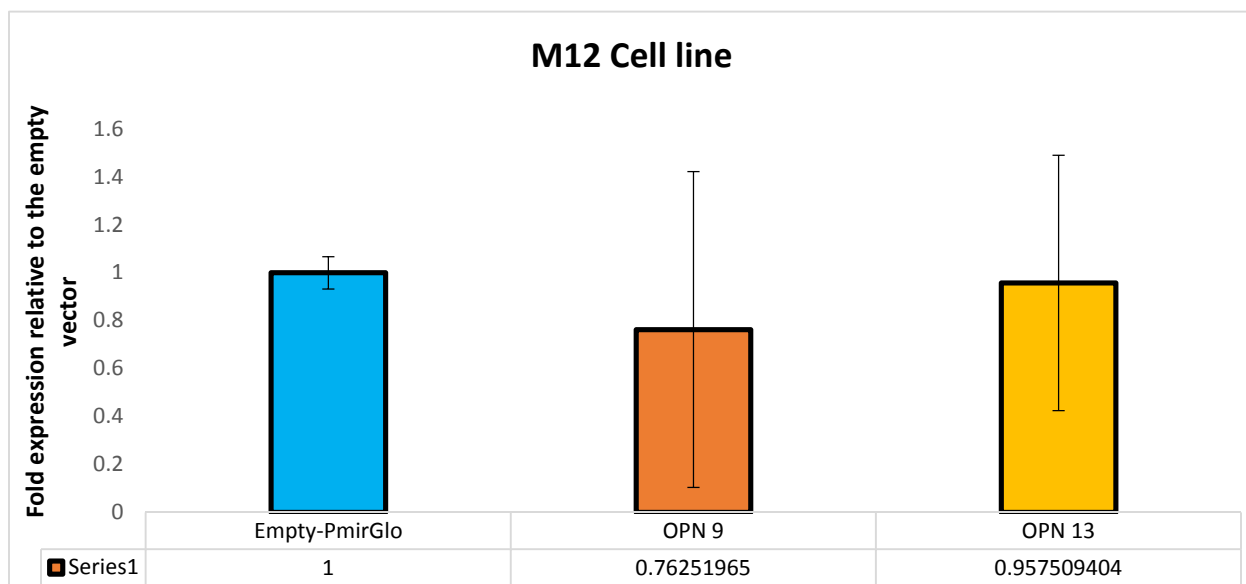


Figure 3-2A: The repression rates of 3'UTR OPN 9 and 3'UTR OPN 13 observed in the M12 cell line. A total of 2.5×10^4 M12 cells per 500uL were plated in a 24-well plate. The firefly luciferase reporter construct containing the 3'UTR OPN 9 or 3'UTR OPN 13 binding seed region (Table 3-1) along with a renilla luciferase plasmid were transfected into the M12 cell line. Reporter activity is presented as the ratio of firefly luciferase normalized to renilla luciferase activity. The expression value for each luciferase:3'UTR construct are reported relative to the empty p-miR-Glo vector ($p \leq 0.05$) with each sample being analyzed in triplicate.

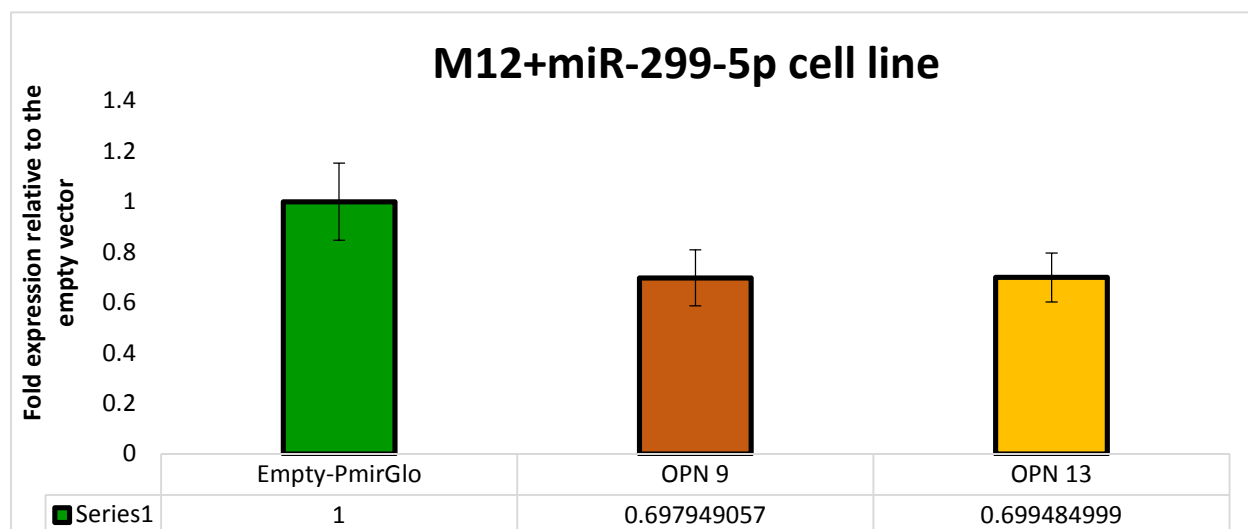


Figure 3-2B: The repression rates of 3'UTR OPN 9 and 3'UTR OPN 13 observed in the M12+miR-299-5p cell line. A total of 2.5×10^4 M12+miR-299-5p cells per 500uL were plated in a 24-well plate. The firefly luciferase reporter construct containing the 3'UTR OPN 9 or 3'UTR OPN 13 binding seed region (Table 3-1) along with a renilla luciferase plasmid were transfected into the M12+miR-299-5p cell line. Reporter activity is presented as the ratio of firefly luciferase normalized to renilla luciferase activity. The expression values for each luciferase:3'UTR construct are reported relative to the empty p-miR-Glo ($p \leq 0.05$) with each sample being analyzed in triplicate.

To confirm that miR-299-5p was truly inhibiting the 3'UTR of OPN, mutants of 3'UTR OPN 9 and 3'UTR OPN 13 were prepared to be transfected into the M12+miR-299 cell line and luciferase assays were conducted. As seen in figure 3-3 the original sequence of the 3'UTR OPN 9 (56nt) and 3'UTR OPN 13 (55nt) were shortened to produce a 25nt and 26nt length 3'UTR OPN 9 and OPN 13, respectively. The new sets of original sequences were further mutated (Figure 3-3). Table 3-2 shows the comparison of mfe between the new sets of original 3'UTR OPN 9 and OPN 13 and their mutants. Figure 3-4 shows that the expression of the mutated 3'UTR' of OPN 9 and 13 in the M12+miR-299-5p had significantly increased 25% and 26% respectively compared to the new set of originals.

3'UTR OPN

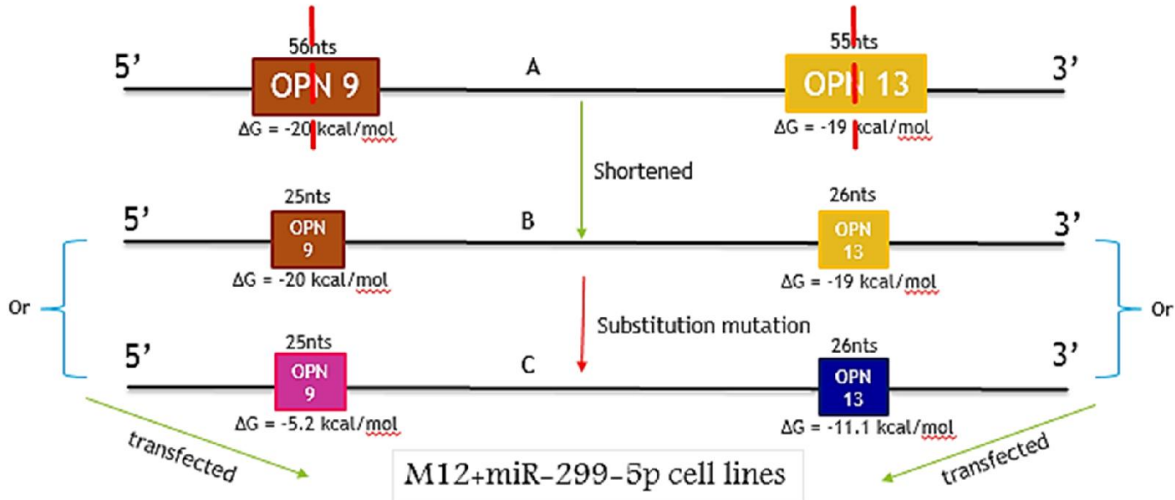
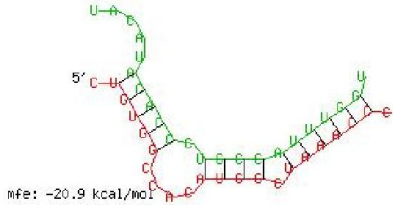
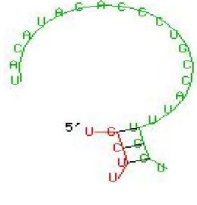
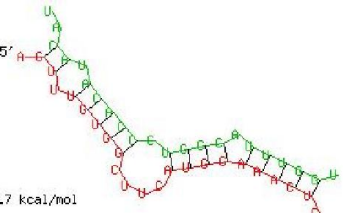
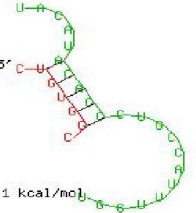


Figure 3-3A-C: The breakdown of old sets of original sequences of 3'UTR OPN 9 and OPN 13 into new sets of original sequences and their mutants. The sequences of the original 3'UTR of OPN 9, OPN 13 and miR-299-5p were submitted to RNA-hybrid [37] program. In order to find the lowest mfe to create mutants, OPN 3'UTR, were shortened to minimize any possible shift in miR-299-5p binding to downstream sites containing a partial match. 3A represents old sets of original 3'UTR OPNs (56nts and 55nts) that were shortened to create 3'UTR OPNs (25nts and 26nts) in 3B. The mfe of both old and new sets of originals were the same, and for both mutants OPN9 and OPN 13 were -5.2 kcal/mol and -11.1 kcal/mol respectively (Table 3-2).

Table 3-2: Comparison of mfe between the new shortened sets of 3'UTR OPN 9 (25nt) and 3'UTR OPN 13 (26nt) and their mutants. Non-mutated and mutated- OPN 9 sequence is in red, OPN 13 is in blue and miR-299-5p is in green. The original 3'UTR mfe's (Table 3-1) are identical to the new sets of shortened sites. As expected the mfe for the mutated sites is considerably reduced for the 3'UTR OPN 9 mutant as -5.2 kcal/mol and 3'UTR OPN 13 as -11.1 kcal/mol.

<p>3'UTR OPN 9 Original MFE: -20.9 kcal/mol (OPN 9)</p> <p>Position 281 Target 5' C CCAC C C 3' UGUGG AUGG UAAACC ACACC UGCC AUUUGG MIRNA 3' UACAUC C U 5'</p>  <p>mfe: -20.9 kcal/mol</p>	<p>3'UTR OPN 9 Mutant MFE: -5.2 kcal/mol</p> <p>Position 281 Target 5' U U 3' GCU UGG MIRNA 3' UACAUCACACCCUGCCA U 5'</p>  <p>mfe: -5.2 kcal/mol</p>
<p>3'UTR OPN 13 Original MFE: -19.7 kcal/mol (OPN 13)</p> <p>Position 1269 Target 5' A U CUUC C 3' GU UGUGG AUGG AAACU CA ACACC UGCC UUUGG miRNA 3' UA U C A U 5'</p>  <p>mfe: -19.7 kcal/mol</p>	<p>3'UTR OPN 13 Mutant MFE: -11.1 kcal/mol</p> <p>Position 1269 Target 5' C C 3' UGUGG ACACC MIRNA 3' UACAUC CUGCCA U UUGGU 5'</p>  <p>mfe: -11.1 kcal/mol</p>

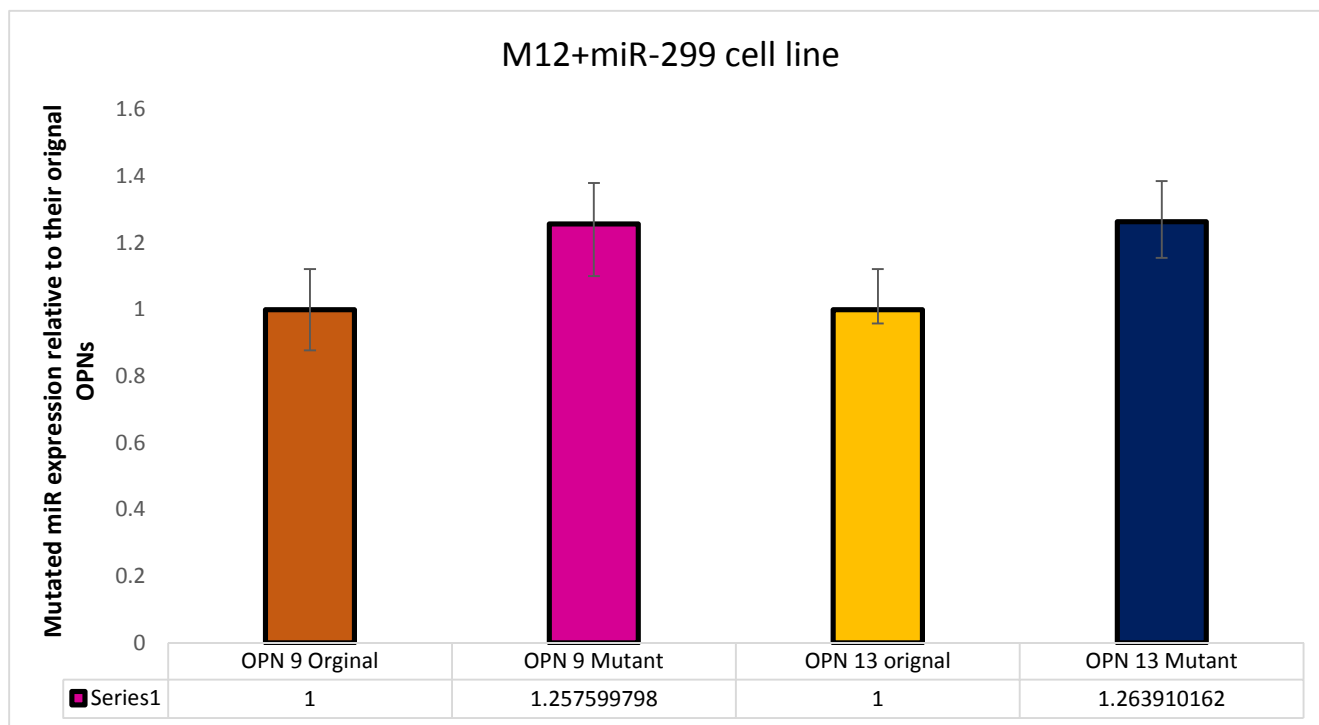


Figure 3-4: The increased expression rates for the mutated OPNs in the M12+miR-299 cell line. A total of 2.5×10^4 M12+miR-299-5p cells per 500uL were plated in a 24-well plate. New sets of original OPNs (brown and yellow bars) and mutated OPNs (purple and blue bars) from Table 3-2 and Figure 3-3 were transfected into the M12+miR-299-5p cell line. The firefly luciferase reporter construct containing the 3'UTR OPN 9 or 3'UTR OPN 13 binding seed region (Table 3-2) or mutated OPN 9 or OPN 13 were transfected together with the renilla luciferase normalization plasmid into the M12+miR-299-5p cell line. Reporter activity is presented as the ratio of firefly luciferase normalized to renilla luciferase activity. Expression values of each luciferase construct is reported relative to the original, non-mutated, OPNs ($p \leq 0.05$) and each sample was analyzed in triplicate.

A proliferation assay was carried out between M12 and M12+miR-299 cell lines measuring growth at 24, 48 and 72 hours in duration. Figure 3-5 depicts that during 24 and 48 hours M12 and M12+miR-299 cells were growing at nearly the same rate with M12+miR-299 cells dividing slightly more often than the M12 cells. By 72 hours the growth of the M12+miR-299 cells started to decline slightly whereas the M12 cell line continued to grow gradually. It was expected that being a tumor suppressor the growth of M12+miR-299-5p will be slower than the M12 cells.

In vitro experiments were conducted to determine if restoration of the potential tumor suppressor miR-299-5p displayed any functional consequences on cell behavior. Since, miR-299-5p overexpression did not affect growth rate, the effect on motility and invasion was measured next. Two sets of experiments were conducted: one with earlier passages of M12 and M12+miR-299 cells, p15 and p18, respectively (Fig. 3-6 A, C) and the other with later passages, p25 and p27, (Fig. 3-6 B, D) respectively. For the migration assay, results varied slightly between an early or late passage, 1.8-fold versus 2.2-fold, respectively (Figure 3-6A compared to Figure 3-6B.). For invasion assay results varied widely between an early or late passage, 1.8-fold versus 0.85-fold, respectively (Figure 3-6C compared to 3-6D).

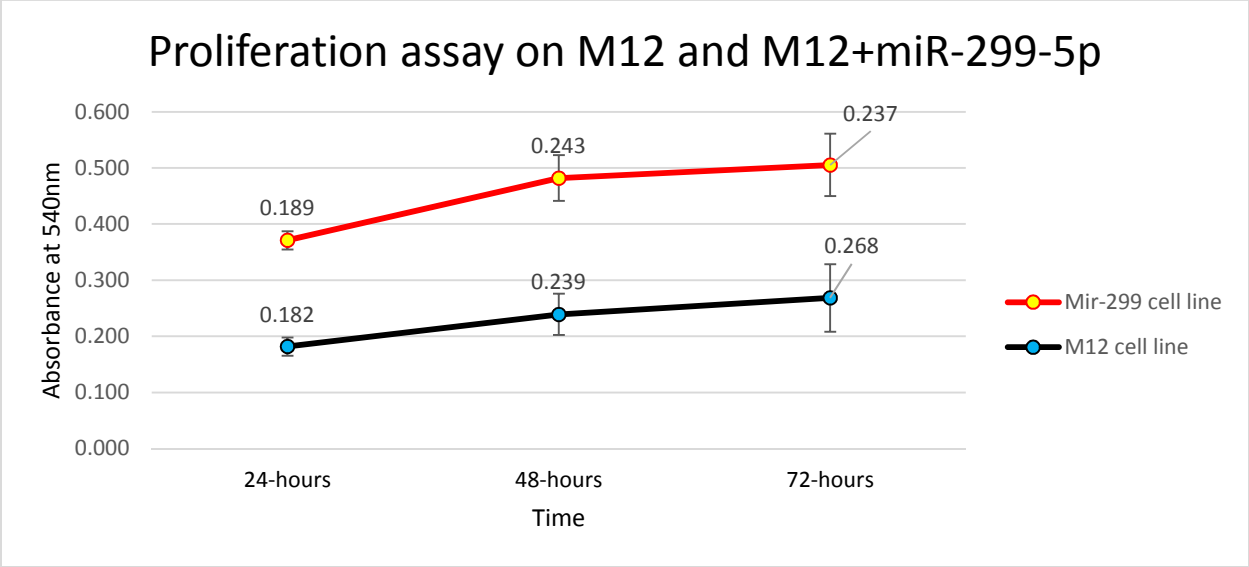


Figure 3-5: Cell proliferation conducted on M12 and M12+miR-299-5p at 24, 48, and 72 hours shows that M12 and M12+miR-299-5p grow nearly at the same rate. A total of 10,000 cell per 100ul were plated in a 96-well plate. As a negative control media without cells were also plated. Each sample was analyzed 9 times, and incubated at 24, 48, and 72 hour intervals. At each point of time 12mM MTT stock solution was added to the wells and incubated at 37°C for 4 hours. After 4 hours, 85uL of the old solution was aspirated off, and 50ul of DMSO was added and mixed thoroughly. The unit was incubated at 37°C for 10 minutes, and the absorbance was measured at 540 nm.

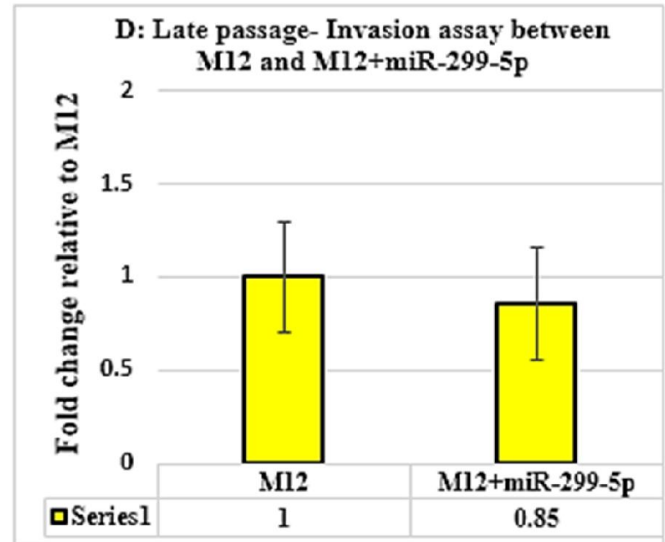
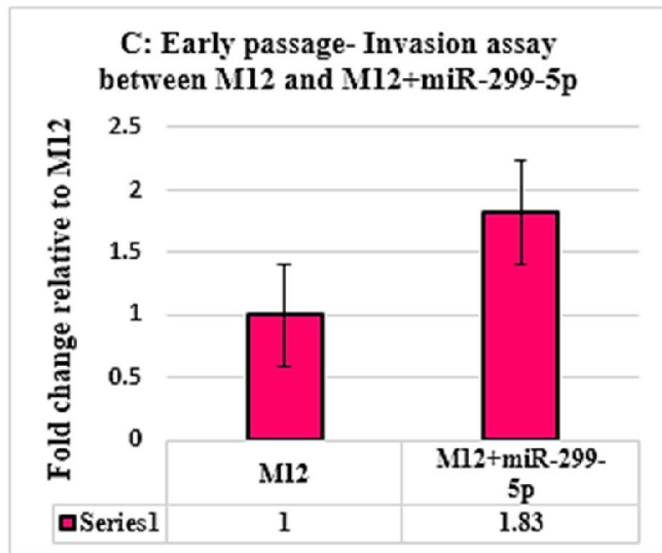
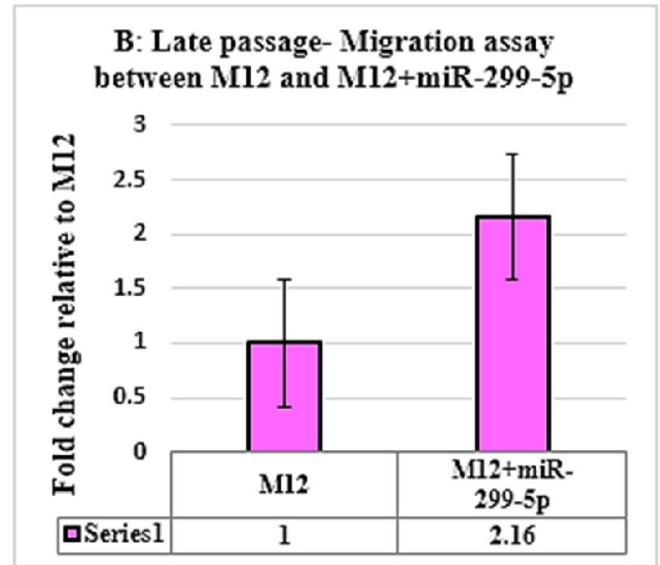
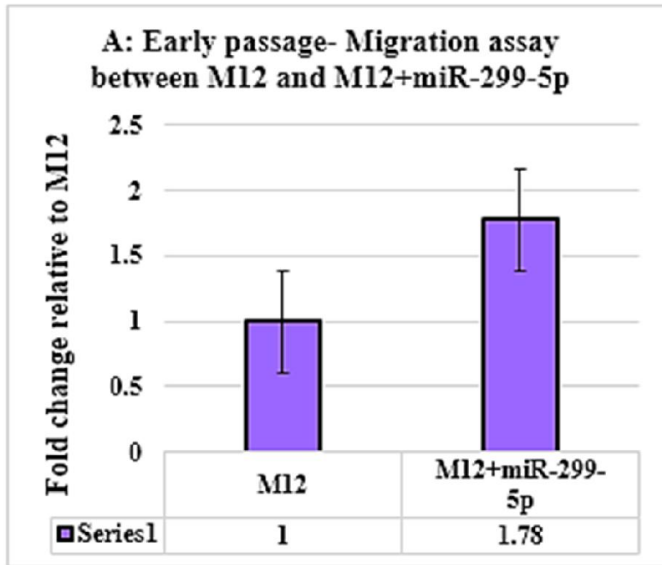


Figure 3-6A-D: A-B Migration assay for early and late cell culture passage of M12 and M12+miR-299-5p cell lines. C-D: Invasion assay for early and late cell culture passage of M12 and M12+miR-299-5p cell lines. M12 and M12+miR-299-5p cells were plated, for migration: 50,000 cells per 200ul; and for invasion: 125,000 cells per 500ul, on a ThinCert™ transwell membrane, with a layer of extracellular matrix added for invasion assay. After 24 hours the membrane containing the cells were mounted on a microscope slide and cells were counted on 5 random fields. Two independent experiments were conducted and each sample was analyzed in triplicate. Each replicate was averaged and expressed as relative to the control (M12 cell line). A-B reports early passaged and late passaged migration assay, respectively. Observing 1.8 and 2.2 fold increase in the movement of cells. C-D reports early passaged and late passaged invasion assay, respectively. Observing, a change of expression between the two passages, 1.8 versus. 0.85.

3.9 Discussion

Previous work compared the expression of miR-299-5p and miR-299-3p in miR array panel screens of RNA extracted from M12 versus parental P69 cell line, confirmatory single miR analysis and in patient samples. Conflicting results were obtained for miR-299-5p where panel analysis suggested it to be a tumor suppressor, no confirmatory results were obtained for single miR analysis and patient samples were split, two suggesting an oncomiR and two a tumor suppressor. Mir array panel screens are notorious for not yielding accurate expression values making single miR analysis imperative. Unfortunately, an error occurred in this PCR reaction so a single miR analysis value was not obtained at that time. In any case it was felt patient data would be more relevant, but as often is the case with patient samples, here too data can vary greatly due to the heterogeneous nature of prostate cancer and the difficulty in accurate staging of a mixed sample. On the other hand, results with miR-299-3p were more consistent. Although the panel miR array screen analysis suggested miR-299-3p might be an oncomiR, single miR analysis and all patient samples tested suggested miR-299-3p as a tumor suppressor.

To resolve these differences and investigate the role of miR-299 in prostate cancer it was decided to restore miR-299 expression in the M12 cell line and investigate the consequences on cell behavior. For this purpose the M12 cell line was stably transformed with a precursor miR-299 mini-gene. To our surprise this resulted in a preferential altered expression of miR299-5p with no increase in miR-299-3p expression and no effect on the expression of two unrelated miRNAs. On the one hand, this unexpected result was favorable as it permitted the analysis of altering only the expression of miR-299-5p alone on cell behavior not complicated by similarly altering the expression of miR299-3p. The reason for the enhanced expression of only miR-299-5p could be

due to diverse nature of miRNAs. They are involved in a wide variety of developmental processes and are essential in networking and regulating gene expression [54]. A specific strand of miRNA is selected by the RISC complex based on the mRNA it aims to target in the cell [55]. The specifics of this selection process is unknown and is currently under study in other laboratories. Our highly metastatic M12 cell line may possess certain mRNAs coupled with a RISC selection process that enables only miR-299-5p to be over produced. This analysis allowed us to proceed with an examination of how the altered expression of only one mature miRNA from a precursor, capable of generating two miRNAs, could by itself affect cell behavior. In retrospect, how the RISC complex accomplishes this feat would be an interesting area of future investigation.

Proliferation, migration, and invasion assays were conducted between the M12 and M12+miR-299-5p cell line to determine the consequence of enhanced miR-299-5p expression on cell behavior. A proliferation assay carried out at 24, 48, and 72 hours' time interval revealed that both cell lines were growing nearly at the same rate, except at 72 hours M12+miR-299-5p declined slightly. In support of this analysis, one study states that miRNAs mapped to the 14q32.31 cluster play an important role in inhibiting cell proliferation in metastatic PCa. Especially, miR-299-5p at 72 hours caused cells to undergo apoptosis [44]. This could be one reason for the decline in cell growth observed in our M12+miR-299-5p cell line at 72 hours. Two sets of migration and invasion assays were carried out for M12 and M12+miR-299-5p cell lines. One set was conducted with earlier passaged cells and the other with later passages. Migration assay results varied only slightly between the two cell lines; however, results varied significantly for the invasion assay between different passage numbers. Cell passaging is defined as number of times cells have either been transferred from vessel-to-vessel [56], or merely, the number of times cells were split and their media changed to ensure healthy living conditions. One study reports that

sequences of changes in miRNA expression occurs during the tissue-culture passage [57]. Cell lines at high passage number experience alterations in growth rates, morphology, response to stimuli, protein expression and transfection efficiency compared to lower passage cells. Passaging cells also causes alteration in cell function such as tumorigenicity, differentiation, gene expression and cell signaling [56]. For our case, miR-299-5p, from the beginning, has shown variability in its nature. For some patient analyses it came out as an oncomiR, while for other patients it came out as a tumor suppressor. The fluctuation in data may have caused miR-299-5p to act in a different way for migratory and invasive processes. Instead of inhibiting, in the early passaged cell line, the M12+miR-299-5p cells migrated like oncomiR cells for both cell migration and invasion. In later passaged cells, the invasion assay was inhibited by miR-299-5p which could explain a shift in cell behavior. As mentioned earlier, cell passage number must have started to alter growth rates and gene expression, causing cells to migrate at a slow rate.

Although miR-299-5p may have behaved opposite to its nature in cell growth and migration, luciferase assays conducted to determine the gene expression of OPN targeted by miR-299-5p confirmed that miR-299-5p was acting as a tumor suppressor. Initially, the luciferase assay was carried out between the two best OPN sites (OPN 9 and OPN 13) within its 3'UTR transfected into both the M12 and M12+miR-299-5p cell lines. The result were favorable and revealed that the expression of OPN's 3'UTR was repressed more in the M12+miR-299-5p cell line than M12 cell line. This suggests that miR-299-5p, stably transformed into the M12 cell line is functional repressing the expression of both OPNs. As mentioned above the study conducted by Shevde *et al.* also showed the expression of OPN to be repressed by co-transfecting miR-299-5p in the MCF10AT (breast cancer) cell line. The results of our analyses were very similar to this published report where in fact 30% repression was observed in our cell lines for both OPN sites while 31%

repression was observed in their study. However, the published study reported repression of only one miR-299-5p binding site in OPN's 3'UTR, whereas we suggested two such binding sites. Furthermore, to confirm miR-299-5p was indeed binding to the 3'UTR of OPN, mutation of the two OPN sites were prepared. To avoid any additional shift downstream of the non-mutated OPNs, the old sets of original OPN were shortened, without disrupting their original seed region for binding to miR-299-5p, hence maintaining the same mfe for both the old and new sets of non-mutated OPN sites. Luciferase assay carried out on the new shortened sets of original OPNs and mutants thereof revealed that while the same repression percentage was observed for the new sets of originals, a higher expression of the mutated OPNs was observed in the M12+miR-299-5p cell line. These analyses confirm that miR-299-5p is a proven target for the 3'UTR of OPN. Often in the literature luciferase assay analyses are conducted on only the wild-type binding sites versus a control vector with no inserted 3'UTR sequences. Here, we report not only luciferase assays conducted on the original OPN binding sites, but also on defined OPN mutations within the binding site for target seed recognition. This further confirms the analyses and ensures that miR-299-5p is one of several miRNAs that binds to the 3'UTR OPN.

In conclusion, from conflicting data on miR-299-5p it is difficult to conclude in which condition this miRNA acts as an oncomiR or as a tumor suppressor. One possibility is that as PCa progresses to later stages of tumorigenesis and ultimately metastasis, miR-299-5p/3p switches roles and acts accordingly. Future experiments will include overexpressing miR-299-3p in the M12 cell line as a miR mimic. The subsequent effect on proliferation migration, and invasion will be measured as done here for miR-299-5p. If miR-299-3p is truly a tumor suppressor then cells should be less migratory and invasive. Ultimately, it will be determined how well

miR-299-3p binds to its proposed target, the androgen receptor by luciferase:3UTR constructs and western blots with the appropriate cell extracts, since antibodies to the AR are available.

On the other hand, it is possible that miR-299-5p and -3p act additively or synergistically with each other thus results obtained with modulating the individual miRNAs may vary widely to what happens *in vitro* where both are expressed perhaps at different levels. To address this issue overexpressing miR-299-5p or -3p within the high or low context of the other maybe required to ultimately determine the consequences of altering miR-299 expression *in vitro*. Considerable work remains to address the importance of miR-299 to the progression of prostate cancer.

Chapter 4

Determine the expression levels of the five selected messenger RNA targets for miR-147b, an oncomiR.

4.1 Introduction:

Hsa-miR-147b is a miRNA that has constantly been found to be up-regulated in previously mentioned (chapter 1), cancer cell lines, human panel analyses, single-miR analyses, and in prostate cancer biopsy tumor samples. However, this miRNA has not been well studied in any of the cancers and there is limited literature that provides information regarding which mRNA(s), miR-147b targets.

The precursor to miR-147b is 80 bases long, and is located on chromosome 15q 21.1 from 45,725,248 bps to 45,725,327 bps, forward strand [59]. Similar to miR-147b is miR-147a (UGUGUG), which differs by only one base in the seed region of miR-147b (UGUGCG) (Table 4-1). Studies have revealed that the expression of miR-147a inhibits cell proliferation and migration [62]. Moreover, miR-147a is involved in a negative feedback loop that inhibits the pro-inflammatory response of macrophages to multiple toll-like receptor (TLR) ligands [60]. A proven target for miR-147a is vascular endothelial growth factor A (VEGFA), but no proven targets for miR-147b have yet been determined [34]. However, a very recent study discovered that miR-147a and miR-147b were overexpressed in left colon cancer compared to right colon tumors. This interesting result suggested that miR-147a and miR-147b could be relevant biomarkers for identifying subtypes of colon cancer [61].

Another miRNA, miR-210 (UGUGCG), contains an identical seed region to miR-147b, but differs greatly for the remainder of the sequence. Based on the 6 nucleotide similarity, the expression of miR-147b and miR-210 could induce similar cellular effects in proliferation, migration and apoptosis. Despite being identical to the hexamer, UGUGCG, the divergent sequence (Table 4-1) between miR-147b and miR-210 appears to play an important role in

determining mRNA targets. Thus, each miRNA has its own unique set of mRNAs that it prefers to bind.

Table 4-1: Sequence similarity between miR-147a, miR-147b, and miR-210. The table contains miRNA sequences, with the seed region highlighted in yellow. The underlined base is where the seed region for miR-147a differs with that of miR-147b and 210.

Micro-RNA	Sequence
Hsa-miR-147a	G <u>UGUGUG</u> GAAAUGCUUCUGC
Hsa-miR-147b	G <u>UGUGCG</u> GAAAUGCUUCUGCUA
Hsa-miR-210	<u>CUGUGCG</u> UGUGACAGCGGCUGA

[Adapted from Seashols, dissertation (2013) [34] and Bertero *et al.* [62]]

Variation in expression due to the divergent sequences of miR-210 and miR-147b was observed in our panel analyses of miRNA expression in the M12 cell line compared to P69 [34]. While miR-210 was found to be overexpressed approximately 2-fold in the M12 cells relative to the P69, miR-147b showed an overexpression of approximately 17-fold. This suggests that major target suppression in M12 occurred due to high overexpression of miR-147b [34].

Previous work conducted in the laboratory by Sarah Seashols, Ph.D., revealed that the stable expression of a miR-147b inhibitor in the M12 cell line, resulted in reduced cellular proliferation, and inhibition of migration, and invasiveness of M12 cells. Moreover, a reduction in subcutaneous tumor growth was also observed in male, athymic nude mice. The Apoptosis-Antagonizing Transcription Factor (AATF), an activator of the cell-cycle inhibitor, p21, was identified as a potential target for miR-147b [34]. These results and the fact that miR-147b is consistently up-regulated in prostate cancer makes it an important miRNA for further study. More importantly, it is necessary to identify additional mRNA targets of miR-147b that could impact the tumorigenicity of M12 cell line as a model for prostate cancer.

A single report was found in the literature dealing with the identification of new miR-147b targets. In this study, different luciferase:3'UTR constructs were co-transfected with a miR-147b mimic into A549 cells, a non-small cell lung adenocarcinoma cell line, and a change in luciferase activity was monitored. Of the 3'UTRs from 15 different mRNAs tested, expression levels of 5 mRNAs, i.e. ALDH5A1, COL4A2, IER5, NDUFA4, and SDHD were reduced upon restoration of miR-147b. Of these 5 targets, NDUFA4 seemed the most promising as it was repressed the most [62].

4.2 A brief overview of the 5 mRNA genes:

ALDH5A1 = aldehyde dehydrogenase 5 Family, Member A1 [63]

ALDH5A1 is a gene located on chromosome 6p22. ALDH proteins play an important role in epithelial cell homeostasis. These ALDH enzymes are present in all cells and in cancer have been shown to protect cells against cytotoxic drugs and oxidative stress. In general, ALDH enzymes are responsible for oxidizing aldehydes to carboxylic acids and they play a critical role in physiological processes like vision, neurotransmission, embryonic development, and ester hydrolysis which serve as binding proteins for androgens and cholesterol [117, 118].

COL4A2 = Collagen alpha-2(IV) chain [63]

COL4A2 encodes six subunits of type IV collagen and is highly conserved across species. It makes up a major structural component of basement membranes, and is associated with cell adhesion and migration. The C-terminal part of the protein, known as canstatin, is an inhibitor of angiogenesis and tumor growth [64]. One study reported that the loss of PDEF an epithelium-specific Ets transcription factor in prostate cancer caused the gain of COL4A2 and its family of genes, hence promoting cell migration and invasiveness [65].

IER5 = Immediate early response 5 [63]

IER5 has been known to be induced by gamma-ray radiation and plays a role in cell death caused by radiation. It has been reported, that IER5 inhibits the transcriptional expression of Cdc25b, which is a co-activator of the androgen receptor, causing proliferation and migration of leukemia progenitor cells to be reduced. Since androgen receptor promotes prostate cancer cell

growth, inhibiting its co-activator may reduce proliferation and migration of prostate cancer cells as well [66].

SDHD = Succinate dehydrogenase complex subunit D [63]

SDHD is a part of mitochondrial subunit of oxidative phosphorylation (OXPHOS) complex II, and is localized in the inner member of mitochondria. Mitochondrial metabolism controls the life of cancer cells and their proliferative properties. Thus, mutation of mitochondrial genes can contribute to tumorigenicity and increase reactive oxygen species (ROS). In fact a variety of mitochondrial DNA and control region mutations have been reported in prostate cancer [67].

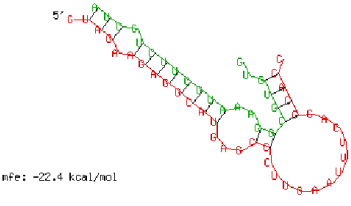
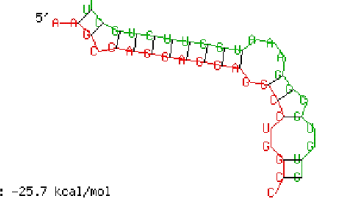
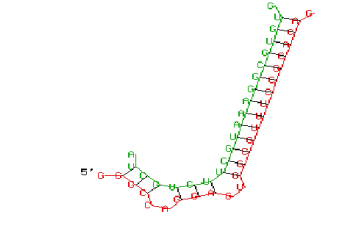
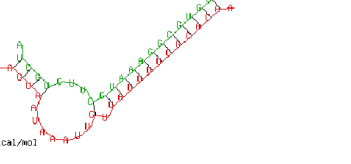
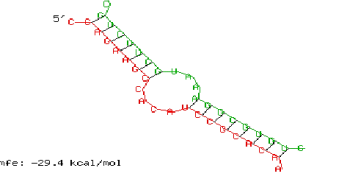
NDUFA4 = NADH dehydrogenase 1 alpha subcomplex 4 [63]

NDUFA4 is an accessory subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase complex I, which does not undergo catalysis. Electrons are transferred from NADH to the respiratory chain. Although a direct role of NDUFA4 has not been found in cancer, the loss of OXPHOS genes, such as NDUFA4 itself, have been shown to contribute to the ability of cancer cells to be self-sufficient and less dependent on OXPHOS [68]. Deficiency in OXPHOS is associated with the decreased availability of ATP, which promotes malignancies and tumor cell expansion [69]. Therefore, it is sufficient to believe that in prostate cancer, part of tumor growth can be held responsible to mutation of mitochondrial genes.

4.3 RNAhybrid analyses for miR-147b and its targets

To assess the binding strength between miR-147b and its 5 mRNA targets, the miR-147b sequence, obtained from miRBase [25], and the 3'UTR sequences of each of the proposed 5 mRNA targets obtained from National Center of Biotechnology Information (NCBI) (Table 4-3) were submitted to analysis via RNAhybrid software [37]. The result of this analyses indicated that miR-147b had a strong binding affinity to all of its mRNA targets (Table 4-2). It will be interesting to determine how well the results of these proposed *in silico* matches correlate with results obtained from expression experiments conducted *in vitro*.

Table 4-2: Binding of miR-147b to its 5 mRNA targets. The sequences of miR-147b and 5 mRNA were submitted to RNAhybrid software [37] to assess their binding strength. The resulting low mfe values, indicate a strong binding between mRNA 3'UTRs and miR-147b.

<p>ALDH5A1</p> <p>MFE: -22.4 kcal/mol position 148</p> <pre> ALDH5A1 5' G A GAG CUUGAAUUUCAG C 3' UAG AGAGGCAU CC GCAC AUC UCUUCGUA GG CGUG miRNA 3' G AA UG 5' </pre>	 <p>mfe: -22.4 kcal/mol</p>
<p>COL4A2</p> <p>MFE: -25.7 kcal/mol position 99</p> <pre> COL4A2 5' A C G UG C 3' AG CAGGAGCA CC C GC UC GUCUUCGU GG G UG miRNA 3' A AAA C UG 5' </pre>	 <p>mfe: -25.7 kcal/mol</p>
<p>IER5</p> <p>MFE: -34.5 kcal/mol position 272</p> <pre> IER5 5' G CA GU G 3' GGC GGA GGCGUUUCCGCACA UCG UCU UCGUAAAGGCGUGU miRNA 3' A G 5' </pre>	 <p>mfe: -34.5 kcal/mol</p>
<p>SDHD</p> <p>MFE: -21.5 kcal/mol position 608</p> <pre> SDHD 5' A AUAUUU U A 3' AGUA G UAUUUUCGCACA UCGU C GUAAAGGCGUGU miRNA 3' A CUU G 5' </pre>	 <p>mfe: -21.5 kcal/mol</p>
<p>NDUFA4</p> <p>MFE: -29.4 kcal/mol position 41</p> <pre> NDUFA4 5' C CACA A 3' CAGAAGC UCCGCACA GUCUUCG AGGCGUGU miRNA 3' AUC UAA G 5' </pre>	 <p>mfe: -29.4 kcal/mol</p>

4.4 The goals of this research

After evaluating the relevance of each mRNA, mentioned above, in prostate cancer and in RNAhybrid software [37]; the goals of this research are as follows:

1. Request the 5 mRNA constructs as luciferase:3'UTR constructs from Bertero, *et al.*, from Institut de Pharmacologie Moléculaire et Cellulaire (IPMC) located in France. Perform standard cloning procedures.
2. Transfect these constructs into the M12 cell line, which naturally expresses miR-147b (17-fold overexpression compared to the parental p69 line) as a model for prostate cancer progression).
3. Perform luciferase assays on these transfection to verify miR-147b: mRNA regulation in M12s as a surrogate for prostate cancer cells. It is expected that luciferase activity will be reduced for mRNAs if truly targeted by miR-147b.

4.5 Methods and Materials:

As previously reported in chapter 2 and 3, cells were cultured, the 5 mRNAs were transfected into the M12 cell line, and luciferase assay were performed.

The only step that differs for this chapter is the:

DNA Synthesis and cloning

Plasmid DNA samples of the 5 mRNA targets, ALDH5A1, COL4A2, IER5, NDUFA4, and SDHD, were obtained from Bertero *et al.* [62] on whatman filter paper. Plasmid DNA came cloned into PsiCHECK-2 vector cut with Xho1 and Not1 sites. Whatman filter paper was cut along the lines, where the plasmid was spotted, and small pieces of the paper were suspended into 50uL of TE or Elution Buffer in an eppendorf tube. The solution was squashed and vortexed vigorously, ensuring elution of the plasmid. The solution was stored at 4°C for a minimum of 30 min and used for transformation into DH5 α *E. coli* cells. Aliquots of 50uL or 100uL of the transformants were plated on agar plates and selected with the drug ampicillin. The forward and reverse primer sequences and length of each of the 5 mRNA's 3'UTR is listed in Table 4-3. The standard miniprep following the manufacturer's protocol (Qiagen) was performed and DNA was sent to VCU-NARF [58] for confirmatory sequencing. Upon confirmation, midiprep DNA (Qiagen) based on the manufacturer's protocol was performed to extract a plasmid suitable for transfection. For transfections only midiprep plasmid products were used due to high purity and concentration of DNA obtained.

Table 4-3: Sequences of the 5 mRNA targets of miR-147b. Accession numbers were obtained from the NCBI database. Length of the 3'UTR indicates the total length of the 3'UTR within which the miR-147b match was found. Forward and reverse primers were obtained from Bertero *et al.* [62].

NCBI - Accession #	Name of mRNA	Length of 3'UTR (bp)	Forward primer	Reverse primer
NM_001080.3	ALDH5A1	400	CGACCTGAGCCTGAGTAAGTGG	AACCCAGAATAACCACAAGAT GTAAGTG
NM_001846.2	COL4A2	288	GGAATTTGCATCCAGCAGCAGCAC	CCTCTGGCCACGGCTGGC
NM_016545.4	IERS5	444	GGCTCATTGGAAGAGGACGATCG	TGCGGGGAAGCAAAGGACCG
NM_003002.3	SDHD	666	GGTGGACAGCCTTCTCTCTTAATC	CATAGAATACATTTTCACATTA GAGATTCCC
NM_002489.3	NDUFA4	420	GAAATGTTTCACTATAACGCTGCTT TAG	CATAGAATACATTTTCACATTA GAGATTCCC

4.6 RESULTS:

Luciferase assays were conducted on transient transfections of each of the 5 mRNA luciferase:3'UTR constructs described above compared to the empty psiCHECK-2 vector, in the M12 cell line, which expresses considerable miR-147b. As anticipated, luciferase activity from each of the constructs was greatly suppressed as compared to the empty vector. The expression of ALDH5A1 was reduced 31%, COL4A2 reduced 59%, IER5 reduced 47%, NDUFA4 reduced 31%, and SDHD reduced 45% (Figure 4-1). Similar assays were performed by Bertero *et al.* [62], in the A549, lung cancer cells yielding the following results: ALDH5A1 reduced 20%, COL4A2 reduced 10%, IER5 reduced 32%, NDUFA4 reduced 41%, and SDHD reduced 5% [62]. Although NDUFA4 was suppressed more in the A549 cells, the majority of these luciferase:3'UTR constructs i.e. ALDH5A1, COL4A2, IER5 and SDHD, were suppressed more in our prostate cancer model M12 cell line than in the A549 cell line. Among the 5 mRNAs, the greatest suppression was observed for COL4A2 and SDHD, followed by IER5, while, the same suppression was observed for NDUFA4 and ALDH5A1.

A comparison between mfe values obtained from RNAhybrid software [37] (Table 4-3), and luciferase activity (Figure 4-1) assay revealed no correlation between these two types of analyses. For instance, the IER showed the most favorable mfe (-34.5 kcal/mol); however, the greatest repression was observed for COL4A2. Similarly, the least favorable mfe was observed for SDHD, but its repression level was greater than either ALDH5A1 or NDUFA4. Although expression of ALDH5A1 and NDUFA4 yielded the same level of repression, their mfes were not alike, i.e., one was -22kcal/mol and other was -29 kcal/mol, respectively (Table 4-4).

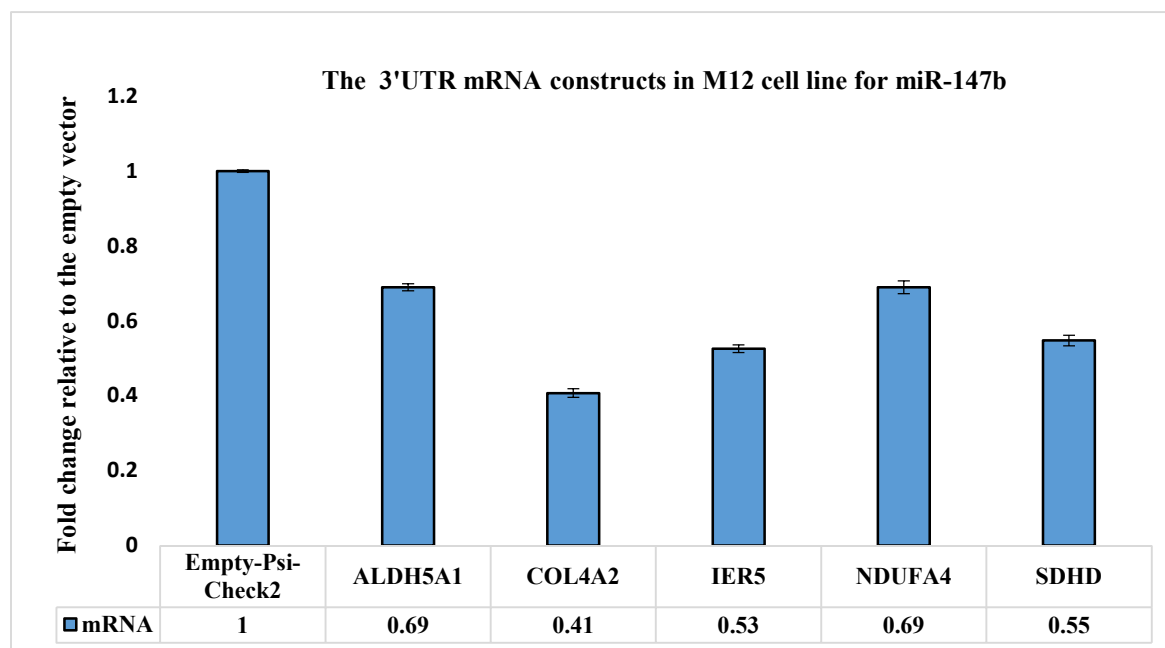


Figure 4-1: Luciferase assay results for transfection of the psiCHECK-2 vector containing the 3'UTR of 5 mRNA targets of miR-147b into the M12 cell line compared to empty vector.

A total of 2.5×10^4 M12 cells per 500uL were plated in the 24-well plate. The 5 mRNA targets of miR-147b were transfected into the M12 cell line, 24-48 hours later cells were lysed, and luciferase activity was measured. The firefly luciferase activity and stop & glo (renilla) were measured at wavelengths of 560 nm and 480 nm, respectively. Firefly luciferase expression is reported normalized to renilla luciferase activity, while fold change values are reported relative to the empty vector. Each mRNA sample was measured in triplicate with a p value ≤ 0.05 . Among the 5 mRNAs, the greatest suppression was observed for COL4A2 followed by IER5 and SDHD at a similar rate, with ALDH5A1 and NDUFA4 yielding similar repression values at the lowest level.

Table 4-4: Results obtained from RNAhybrid software, and luciferase assay. A clear correlation between the two analyses is not observed. The lowest mfe was observed for IER5, however the highest repression was observed for COL4A2. Similarly, ALDH5A1 and NDUFA4 repressed at the same level but their mfes differed.

mRNA	Mfe from RNAhybrid (kcal/mol)	Repression percentage (%)
ALDH5A1	-22.4	31
COL4A2	-25.7	60
IER5	-34.5	47
SDHD	-21.5	45
NDUFA4	-29.4	31

4.7 Discussion:

MiR-147b appears to function as an oncomiR in prostate cancer. Its high level of expression has been consistently observed in every type of analyses the Zehner laboratory has conducted [34]. Human panel miR arrays, single-miR analysis, and prostate cancer biopsy samples have confirmed miR-147b to be an oncomiR. Reduced cell proliferation, migration and invasion were observed when miR-147b was inhibited in the M12 cell line [34]. Thus an investigation into those targets driving a tumorigenic outcome is important. At the time there was a limited literature on proven miR-147b targets and especially not in prostate cancer. Although some bioinformatic tools had suggested potential targets for miR-147b, there was no research, to validate a regulatory role. Despite limited sources, Sarah Seashols, Ph.D., was able to identify a new target for miR-147b as AATF.

Each miRNA is proposed to target hundreds of mRNAs and probably cancer results from the targeting of several mRNAs not just a change in expression of a single mRNA. Thus, it became important to identify other mRNA targets whose change in expression via miR-147b regulation could contribute to prostate cancer progression. .

A recent paper disclosed 5 new mRNA targets, ALDH5A1, COL4A2, IER5, SDHD, and NDUFA4 as potential targets for miR-147b in the A549 lung cancer cell line [62]. Here we show these same 5 mRNAs are relevant targets for miR147b regulation in prostate cancer as well, but not necessarily at the same level. Some mRNAs were suppressed more than others and did not exhibit a direct correlation to those results obtained in the A459 lung cancer cell line. Because miRNAs are tissue specific in their expression, it is expected that their resulting effect on target regulation will differ in different cell lines as seen here. One reason for this can be the selective

presence or lack of other miRNAs in the M12 cell line, which compete with mir -147b for target binding. In the latter case, the lack of competitive miRNAs could favor the mir147b binding and yield a greater repression of the mRNA targets.

Transfection of the 3'UTRs of ALDH5A1, COL4A2, NDUFA4, SDHD, and IER5, in the M12 cell line showed a reduction in the luciferase activity. Most repression in the activity was observed by COL4A2, followed by IER5, SDHD, NDUFA4, and ALDH5A1. SDHD, NDUFA4, and ALDH5A1 are mitochondrial enzymes that are crucial for normal tricarboxylic acid cycle and electron transport chain (ETC) activity. NDUFA4 is a subunit of ETC complex I and SDHD is a subunit of ETC complex II [116]. COL4A2 is associated with cell adhesion and migration [65], and IER5, is induced by gamma-ray radiation, which causes cell death by radiation [66]. Down-regulation of these mRNAs indicates that miR-147b alters mitochondrial and apoptotic function in the M12 cell line, which facilitates the progression of tumor. To support this analyses, a similar study was performed by Grosso *et al.* with miR-210 in the A549, lung cancer cell line. MiR-210 was found to be up-regulated in this cell line, and luciferase assay showed that miR-210 inhibited the expression of COL4A2, ALDH5A1, NDUFA4, and SDHD, confirming an alteration to the mitochondrial respiration in the A549 cells. To further confirm their analyses, electron microscopy performed on the cells containing miR-210, revealed a re-arrangement of cristae in the mitochondria compared to the normal mitochondria morphology [70]. In our human panel analyses both miR-147b and miR-210 were up-regulated, and classified as oncomiRs. As mentioned above, these miRNAs share a common seed region, hence they target some of the same mRNAs. These miRNAs also induce similar cellular effects in term of proliferation, migration and apoptosis [62]. Similarly just like miR-210, our luciferase assay also confirms that miR-147b modifies normal

mitochondrial and apoptotic behavior of cells, and promotes tumorigenesis. Overall, these luciferase:3'UTR assay results validate a direct interaction between miR-147b and all its targets.

RNAhybrid results did not completely correlate with the luciferase assay results. The lowest mfe was observed for IER5, however the most repressed luciferase activity was observed for COL4A2. The highest mfe was observed for SDHD, but it repressed better than ALDH5A1 or NDUFA4. Even after observing the same repression rates for ALDH5A1 and NDUFA4, their mfe's differed, in fact NDUFA4 had a lower mfe than ALDH5A1. Although RNAhybrid [37] is a useful tool for predicting the binding strength of a miRNA to mRNA interaction, this software does not take into account the cell lines in which the repression activity will be measured. MiRNAs are cell specific, hence our M12 cell line may contain other miRNAs that may have a better binding affinity to the 3'UTR of the 5 mRNA targets, than miR-147b. On the other hand, they may lack other competing miRNAs with a better binding affinity to any one of these targets, giving miR147b a regulatory advantage not duplicated in the A459 cell line. Overall, RNAhybrid is a great tool to predict an interaction between the miRNA and its target mRNA, but experimental validation, is required to validate such an interaction and the consequences thereof to cell behavior.

In conclusion, the results of luciferase assay confirm a direct binding between all 5 mRNAs and their target, miR-147b. However, the computational analyses did not confer with luciferase assay analyses. Conveying, that miRNAs are specific to each cell line, and their behavior may not always adhere to the behavior observed in computational analyses.

Chapter 5

A bioinformatics approach to identify microRNAs present in the body fluids of normal subjects.

5.1 Introduction:-

The discovery of miRNAs is an emerging field in cancer research with potential novel techniques for diagnoses and therapy [71]. MiRNAs are small regulatory RNA molecules that regulate the activity of specific mRNA targets and play an essential role in physiological and pathological processes [72], such as gene-expression and aberrant expression in the pathogenesis of cancer, cardiac, immune-related, and other diseases [73]. While research on identifying tumor specific miRNA has been ongoing for more than a decade now, miRNAs have also been discovered in serum, plasma, and other body fluids [73]. A study performed by Chen et al. [75] demonstrated by using a Solexa Sequencing platform that by extracting RNA from the serum, they could identify unique miRNA expression profiles, for lung cancer, colorectal cancer, and diabetes patients compared with healthy subjects. This study also revealed that miRNA expression profiles differed between serum and blood cells of lung cancer patients, while similar miRNA expression profiles were seen in the serum and blood cells of healthy subjects. Overexpression of specific miRNAs within a tumor can be observed in body fluids of patients; therefore, involvement of miRNAs in early tumorigenesis, tumor progression and metastasis could serve as diagnostic or prognostic markers for a variety of cancers [71]. Novel or improved approaches for detecting common epithelial malignancies are required to reduce major health issues caused by cancer. Although conventional strategies for blood-based biomarker discovery using proteomic technologies have shown promise, the development of clinical methods to identify biomarkers, for early tumor detection, in cancer still requires improvement [74, 75].

5.2 Stability of miRNAs:

Decades ago when miRNAs were newly identified, many investigators predicted that extracellular RNA could not survive in blood due to presence of potent ribonucleases [97]. However, slowly this misconception faded as additional studies revealed that circulating extracellular RNA in serum is protected from plasma RNase activity [98]. Chen et al [75] showed that serum and plasma contain a large number of miRNAs and these miRNAs remain stable after being subjected to severe conditions such as boiling, very low or high pH, extended storage, and 10 freeze-thaw cycles that would normally degrade most RNAs. Many theories have attempted to explain the possible mechanisms responsible for such high quality stability. One theory suggests that RNA present in plasma is protected by inclusion in lipid or lipoprotein complexes [97]. Other reports have suggested the secretion of mRNA and miRNAs across cell membranes either in microvesicles or in small membrane vesicles of endocytic origin called exosomes [99, 100]. A study performed by Taylor et al demonstrated that miRNA contained in tumor exosomes are functional and can suppress the mRNA that encodes signal transduction components within T-cells. [101]. Skog et al, demonstrated that exosomes secreted by glioblastoma cells containing mRNA, miRNA and angiogenic proteins are taken up by normal brain microvascular endothelial recipient cells. The messages delivered by tumor-derived exosomes are further translated by these recipient cells to stimulate proliferation of a human glioma cell line [102]. Prior reports indicate that cancer patients have elevated levels of tumor-derived exosomes in plasma compared to non-cancer controls. Although normal cells within peripheral circulation can contribute to exosome population, the primary source of circulating exosomes in cancer patients is thought to be from the tumor [103]. Although the presence of miRNA in exosomes can explain their stability in serum, other reasons for stability can be due to protection by chemical processes or associations with protein complexes.

Currently there is a lack of an established endogenous miRNA control for normalization in plasma or serum. The most common control used to normalize miRNA for qRT-PCR data is U6 small nuclear RNA (RNU6B); however, this RNA was found to be less stably expressed than miR-93, miR-106a, miR-17-5p, or miR-25 in serum. [104]. Thus additional studies are necessary for the identification of an accurate normalization control for each type of cell, tissue or body fluids to be studied.

5.3 Identifying miRNAs in body fluids:

An important aspect of miRNAs is the possibility that these molecules are secreted, circulate and subsequently affect cellular behaviors elsewhere in the body producing favorable conditions for disease progression. In 2008 Lawrie *et al.* introduced miRNAs as a new class of biomarkers in serum. In this study miRNAs were purified from patient serum using the Trizol reagent and certain miRNAs were quantified using Taqman based PCR. Results showed that the expression pattern of miR-155, miR-210 and miR-21 distinguished large B-cell and lymphoma patients from normal people [80]. Several other studies have shown miRNA expression profiles related to tumor classification. Mitchell et al found that by measuring serum levels of miR-141, they could distinguish patients with prostate cancer from healthy subjects [74]. Similarly Chen et al demonstrated that levels of miR-25 and miR-223 were more elevated in cancer patients than those of healthy individuals [75]. Ng et al showed that miR-92 was more significantly increased in colorectal cancer than in gastric cancer or normal subjects [105]. Wong et al reported that miR-184 showed significantly higher expression in tongue squamous cell carcinoma cells than in paired normal cells [106]. While these studies identified miRNAs expression level in plasma or serum, one report evaluated miRNA expression levels in urine, breast milk, tears, saliva, cerebrospinal fluid, seminal fluid, amniotic fluid and other fluids. These body fluids were obtained from women

in different stages of pregnancy or in patients with different urothelial cancers. The results of this study demonstrated that miRNAs were present in all body fluids, several of the abundant miRNAs were common among multiple fluid types while others were enriched in specific fluids. The overall conclusion of this study was that although miRNAs were ubiquitous to all biological fluids, specific compositions components at different concentrations were observed for each body fluid. These findings are important because it presents a useful correlation between specific miRNAs in body fluids and various diseases. Hence, this study provides the insight that miRNAs could serve as useful biomarkers in clinical usage [76].

5.4 Aim of this research:

The purpose of this research is to identify miRNAs present in six body fluids of healthy male and female volunteers: venous blood, menstrual blood, vaginal fluid, semen, saliva, and feces through a next-generation sequencing approach. In our case, such miRNAs will be first identified in normal healthy human body fluids. The result of this will not only help us determine prominent miRNAs in body fluids that could serve as biomarkers, but miRNAs that can act as positive controls against which the content of prostate cancer patient body fluids can be compared. It is essential to determine which miRNAs are found normally in body fluids attained from healthy people; so that in analyzing prostate cancer patients' body fluids, the loss or gain of miRNAs could be diagnostic of disease.

The advantage of using RNA sequencing versus miR panel array analysis is that by the former all possible, including novel, miRNAs will be identified. Screening of microarray panels are only able to yield data on those already known and included miRNAs. The promise of RNA sequencing is that all miRNAs both known and unknown can be deduced and quantitated.

Moreover, the analysis of microarray panels has already been done in the literature, but RNA sequencing of all body fluids relevant to forensic science or to prostate cancer has not yet been reported. Hence, we propose that this is the more encompassing study that should be done.

Previous laboratory work performed by students within the Forensics Science and Biochemistry Departments at VCU-Medical Center under the guidance of Drs. Sarah Seashols and Zendra Zehner, included extracting RNA from 4 different donors for each of the body fluids analyzed. After evaluating quality and quantity of the acquired RNA using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA [34]) and NanoDrop ND-2000 Spectrophotometer (Thermo-Fisher Scientific, Inc; Waltham, MA [34]), the required concentration of RNA (100ng/uL), was sent off to the VCU-NARF core [58] for RNA sequencing using the Illumina-HiSeq platform. The Illumina-HiSeq, RNA sequence data (RNA-SEQ), was analyzed using the CLC Genomics workbench version 7.0.4 and annotated against miRBase (v.20, 2013) [13].

5.5 The CLC genomics workbench:

CLC is a high-throughput genomic software that supports features such as *de novo* assembly, where it offers comprehensive support for a variety of data formats such as the 454 FLX system, Illumina Genome Analyzer, Illumina HiSeq, HiScan, MiSeq sequencing systems, SOLiD system, and Ion Torrent system. Mapping of both short, long and paired reads can be done in this software. RNA-SEQ analyses, in which we are interested is, based on reference genome or mRNA sequencing reads. This workbench is able to calculate expression levels and determine novel exons or miRNAs. It provides core tools to trim the adapter sequences on next generation sequencing data, provides transcriptomic analyses such as extract and count, common occurring sequences, and annotate and merge data against the desired database or reference genome [79].

5.6 Methods to curate RNA-SEQ data in the CLC genomic workbench:

A total of 6 body fluids were analyzed: venous blood, menstrual blood, vaginal fluid, semen, saliva, and feces. Four samples per body fluid were sent off for Illumina-HiSeq sequencing; thus a total of 24 data files were obtained for analyses. In general for Illumina-HiSeq sequencing, a population of RNA is first converted into a library of cDNA fragments with adapter sequences attached to one or both ends. Through amplification process each molecule is sequenced in a high-throughput manner to obtain sequences from either, one end (single-end sequencing) or both ends (pair-end sequencing). Sequences generated are usually 70-200bp long in length [77]. In our case only single-end sequencing data was generated, due to the panel size of the cDNA template and each sequence in the RNA-SEQ data consisted of a 5' primer sequence, 5' adapter sequence, miRNA sequence, 3'adapter sequence, barcode, and 3' primer sequence (Figure 5-1).



Figure 5-1: The final PCR product and the location of each adapter sequence. Once RNA has been sequenced the final PCR product consists of the following, a 5' small RNA sequencing primer, 5' adapter sequence (green), miRNA insert (blue), 3' adapter sequence (purple), barcode (black), and multiplex index read sequencing primer.

[The figure idea adapted from the New England BioLabs manual [107]]

To annotate the RNA-SEQ data, three major steps in the CLC genomic software were applied: (i) Trimming off the adapter sequences, (ii) extract and count sequences that occurred more than once and (iii) annotate against a miRNA database, for our purpose miRBase (v.20, 2013) database was used [13].

(i) Trimming off the adapter sequences:

Fastq formatted files were imported into the CLC genomic software, single-end-sequence option was selected and NCBI/Sanger or Illumina 1.8 pipeline was used. Before trimming off the adapter sequences a trim adapter sequence file was created, where 5' adapter sequence (5'GTTTCAGAGTTCTACAGTCCGACGATC 3'), reverse complimented 3' adapter sequences (5'AGATCGGAAGAGCACACGTCT 3') were recorded and plus and minus strands were selected, respectively. For each adapter sequence the remove adapter option was selected. This option removes adapter sequences and all the nucleotides 5' of the match, while preserving everything 3' end of the match. The same action was applied for the 3' adapter sequence, where together with the adapter sequence all the nucleotides 3' of the match were removed and everything 5' end of the match was preserved. The alignment score [14] for each of the adapter sequence was set as: - mismatches = 2, gap-cost = 3, minimum score for matches found within the read was set to 15, and minimum score for end matches was set to 4. Examples of how mismatches and gap-costs were calculated are shown in figure 5-2 [79]. Reads were trimmed based on quality score of 0.05 and maximum of 2 ambiguous nucleotides were removed. Reads below length 15 and above length 1000 were discarded.

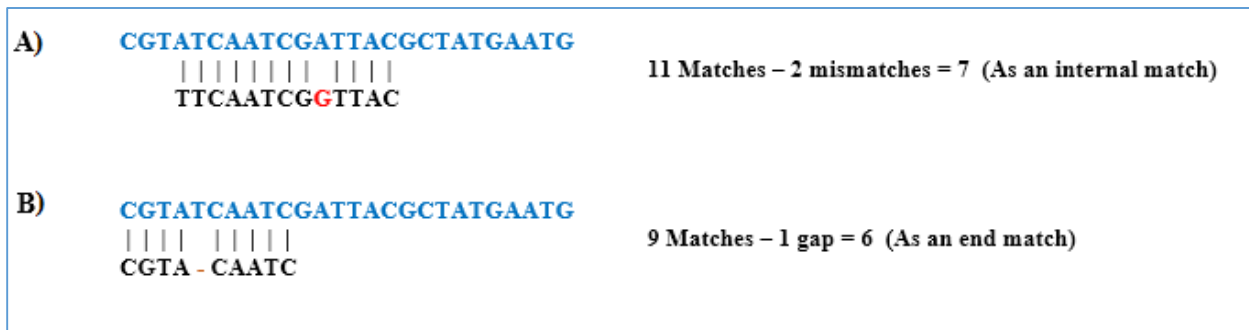


Figure 5-2: Examples showing how mismatches and gap-costs are calculated based on Smith-Waterman algorithm. A list of adapter sequence in the CLC genomic software has to be created in order to perform the trimming step on the reads from RNA-SEQ data. When a match on the read is found a set of scoring thresholds are adjusted for each adapter sequence. The first step includes setting the costs of mismatches and gaps. According to the Smith-Waterman algorithm [78] a match is rewarded one point, this score cannot be changed. A mismatch costs 2 points and a gap, such as insertions or deletions, costs 3 points. Mismatch based on internal matches: for example A) a total of 11 matches was found ($11 \times 1 = 11$), with 2 mismatches ($2 \times 2 = 4$), hence the alignment score equals to $11 - 4 = 7$. Mismatches found on end indicates a match found on either end of the read. Example B) shows that 9 matches were found with 1 gap ($1 \times 3 = 3$) hence the total alignment score was 6 for the end match.

[The figure idea adapted from the CLC genomic software manual [79]]

(ii) Extract and count trimmed reads:

This step can be performed either on the original (non-trimmed) fastq file or trimmed file generated from the trimming off the adapter step. If conducting analyses on the original file, the adapter trimming step is selected, else if analyzing the trimmed file, that option is left blank. Reads below length 15 and above length 55 were discarded. Minimum sample count was equaled to 1. Minimum sampling counts is number of small RNAs tags needed to count the reads. Counting of the re-occurring reads is based on perfect matches found between sequences. Therefore, a threshold of 1 indicates a lot of unique tags are created, while high threshold becomes selective and only few tags are generated.

(iii) Annotation and merge counts:

The file generated from the extract and count step is used for the annotation and merge count step. Before this step was performed, miRbase release 20 database was downloaded in the CLC genomic software. Together with miRbase v.20, other miRNA database can also be downloaded and selected. *H. sapiens* species was selected as the organism; however more than one species can also be selected. The order in which the species are selected is important because the tags annotate iteratively based on the order specified. Next, the parameters for alignment of the tags against the annotation sources were set as: additional upstream bases = 2; additional downstream bases = 2; missing upstream bases = 2; and missing downstream bases = 2; these parameters are set as default by the software for only miRbase annotation source. Additional upstream and downstream bases defines how many bases the tag is allowed to extend the annotated mature region at 5' end or 3'end and can still be identified as mature. Missing upstream and downstream bases indicate how many bases a tag is allowed to miss either 5' end or 3' end compare

to the annotated mature region and still be identified as mature. Maximum mismatches was set to 2. Next, the types of file to be generated were selected.

5.7 RESULTS:

The types of files selected to be generated from the annotation and merge count steps were as follows: a) Annotated file- which contained information such as small RNA sequence, length of those sequences, count, name of the miRNA, match type as in whether the sequence was qualified as mature 5', mature 3', mature 5' or 3' sub, mature 5' or 3' sub variant, and mature 5' or 3' super or super variant and mismatches; b) annotated grouped sample- this is a file where all the mature 5' or 3', sub/sub variants/super/super variants are grouped together and total count is shown; and a summary report where a summary of how many sequences were annotated and the percentages were reported. Tables 5-1 A-F represents the data generated from trimming off the adapter sequence step. For the most part 74-99% of the adapter sequences were removed, indicating that the majority of the adapter sequences were successfully trimmed off.

Table 5-1A: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for venous blood.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
V.Blood_7309	1,957,364	91.7	1,767,370	24.6	90.29%
V.Blood_7311	1,737,441	90.8	1,578,616	23.9	90.86%
V.Blood_7318	2,381,563	90.9	2,134,571	24.2	89.63%
V.Blood_7319	2,014,376	90.8	1,807,780	24.9	89.74%

Table 5-1B: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for menstrual blood.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
MB_7329_Day2	595,548	97.5	561,540	31.4	94.29%
MB_7318_Day3	1,039,921	93.0	925,872	27.8	89.03%
MB_7315_Day6	1,983,903	94.5	1,771,937	30.2	89.32%

Table 5-1C: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for vaginal fluid.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
VF_7314	476,369	99.6	467,108	34.1	98.06%
VF_7315	563,933	93.5	481,684	29.6	85.42%
VF_7319	552,911	92.0	467,583	27.9	84.57%
VF_7321	3,511,331	97.8	3,353,973	32.4	95.52%

Table 5-1D: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for semen.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
Semen_7311	1,170,387	98.1	1,117,742	31.3	95.5%
Semen_7322	362,993	97.7	349,011	31.6	96.15%
Semen_7520	369,643	98.1	366,434	31.9	99.13%
Semen_9602	482,072	97.8	457,482	31.8	94.9%

Table 5-1E: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for saliva.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
Saliva_7319	1,501,694	95.0	1,351,059	30.1	89.97%
Saliva_7321	1,055,656	95.0	1,006,329	29.6	95.33%
Saliva_7322	1,379,499	99.3	1,348,952	33.7	97.79%
Saliva_8425	1,986,277	98.4	1,931,595	32.6	97.25%

Table 5-1F: Results of trimming off the adapter sequences from reads contained in the RNA-SEQ data obtained for feces.

Body Fluid	Total number of reads	Average length before trimming	Number of reads after trim	Average length after trim	Percentage trimmed
Feces_7318	1,717,555	91.3	1,363,578	27.8	79.39%
Feces_7319	1,452,225	93.5	1,265,847	28.9	87.17%
Feces_8401	1,420,982	96.3	1,287,695	32.0	90.62%
Feces_8407	2,471,044	88.5	1,831,558	25.1	74.12%

The annotation results varied widely among each body fluid, blood annotated 78%, menstrual blood annotated 4.4%, vaginal fluid annotated 1.3%, semen annotated 8.8%, saliva annotated 1.3%, and feces annotated only 0.20% (Table 5-2).

Table 5-2: The percentage each body fluid was annotated against the miRBase database. The four files per body fluid were averaged together. Thus values are reported here as an average of total reads after the extract and count step and an average of total reads annotated.

Body Fluid	Average number of reads after extract and count	Average number of reads annotated against miRBase	Percent annotated (%)
Venous blood	1,703,867	1,321,778	77.58
Menstrual blood	1,027,420	45,095	4.39
Vaginal fluid	1,183,853	15,295	1.29
Semen	568,640	49,781	8.75
Saliva	1,398,650	18,377	1.31
Feces	1,422,408	2,804	0.20

Aforementioned, the annotated grouped file contained miRNAs that were classified as either mature 5' or mature 3'. When RNAase III enzyme Dicer with its cofactor TRBP, snips off the loop of the pre-miRNA, a double stranded RNA duplex approximately 18-24 nucleotides in length is released. This double stranded RNA is further incorporated into RNA-induced silencing complex (RISC) to help retain a mature miRNA. Depending on the tissue, cell and the mRNA a miRNA targets, either a mature 5' strand or mature 3' strand is produced. Similarly, CLC genomic workbench, based on the types of cells found in each body fluid, classified each miRNA as either mature 5' or mature 3'. Henceforth, the annotated data, represented as a list of abundant and unique miRNAs found in each body fluid (Table 5-4), and graphs representing common miRNAs found within each body fluid (Figure 5-3), was also organized in a similar manner.

Table 5-4 and figure 5-3 A-L reported below show that there are many common miRNAs found in the body fluids analyzed for this research. To name some of the most common mature 5' miRNA to occur within all body fluids is miR-21. Second most common miRNA to occur in all body fluids except semen is miR-451a. MiR-486 and miR-486-2 have always occurred together in venous blood, menstrual blood and feces. While miR-26a-1 and miR-26a-2 have also occurred together in blood, semen, and feces. MiR-30d was also present in all body fluids except menstrual blood and vaginal fluid. MiR-99a was found in menstrual blood, vaginal fluid, and semen, while miR-205 was found in menstrual blood, venous blood, and saliva. To name some of the common mature 3', miR-101-1, miR-101-2, and miR-22 occurred in all body fluids. MiR-148a occurred in all body fluids except venous blood. MiRs-200a, 200b and 200c occurred together in menstrual blood, vaginal fluid, and semen. MiR-103a-1 was found in venous blood, menstrual blood and feces. MiR-203a was found in all body fluids except venous blood, and

semen. MiR-27b was found in all except venous blood and feces. MiRs-363, 92a-1, 92a-2 were found in venous blood and feces, except miR-363 was also found in semen. A compiled version of miRNA found in body fluid is listed in table 5-3.

Table 5-3: Compiled version of miRNAs found commonly in two or more body fluids.

VB = venous blood, MB = menstrual blood, and VF = vaginal fluid.

MiRNA mature 5'	Body Fluid
MiR-451a	VB, MB, VF, saliva, and feces
MiR-486	VB, MB, and feces
MiR-26a-1, Mir-26a-2	VB, semen, and feces
MiR-30d	VB, semen, saliva and feces
MiR-99a	MB, VF, and semen
MiR-205	MB, VB, and saliva
MiR-21	All body fluids
MiRNA mature 3'	Body Fluid
MiR-101-1, miR-101-2, miR-22	All body fluids
MiR-148a	MB, VF, semen, saliva, and feces
MiR-200a, 200b, and 200c	MB, VF, and semen
MiR-103a-1	VB, MB and feces
MiR-203a	MB, VF, saliva, and feces
MiR-27b	MB, VF, saliva, and semen
MiR-92a-1, 92a-2	VB, and feces
MiR-363	VB, feces, and semen

Table 5-4 Annotation results of each body fluid. Reporting top mature 5' and 3' miRNAs found abundant in each body fluid. The ones in bold are unique miRNAs, found among the top 10-20 miRNAs in all body fluids. Unique miRNAs were found through a UNIX command. The count (not shown) of miRNA for each body fluid is listed in descending order.

Body Fluid	Top mature 5' miRNAs	Top mature 3' miRNAs
Venous blood	<ul style="list-style-type: none"> • miR-451a • let-7g • let-7i • miR-486-2 • miR-486 • miR-21 • let-7f-2 • let-7f-1 • miR-191 • miR-185 • miR-20a • miR-30e • miR-26a-1 • miR-26a-2 • miR-30d • let-7a-2 • let-7a-1 	<ul style="list-style-type: none"> • miR-144 • miR-101-1 • miR-101-2 • miR-92a-1 • miR-92a-2 • miR-25 • miR-22 • miR-126 • miR-103a-1 • miR-103a-2 • miR-363
Menstrual blood	<ul style="list-style-type: none"> • miR-21 • miR-451a • miR-205 • miR-99a • let-7f-2 • let-7g • let-7f-1 • let-7i • miR-486-2 • miR-486 • miR-146b 	<ul style="list-style-type: none"> • miR-203a • miR-148a • miR-200b • miR-27b • miR-22 • miR-200c • miR-200a • miR-101-2 • miR-101-1 • miR-103a-1 • miR-143

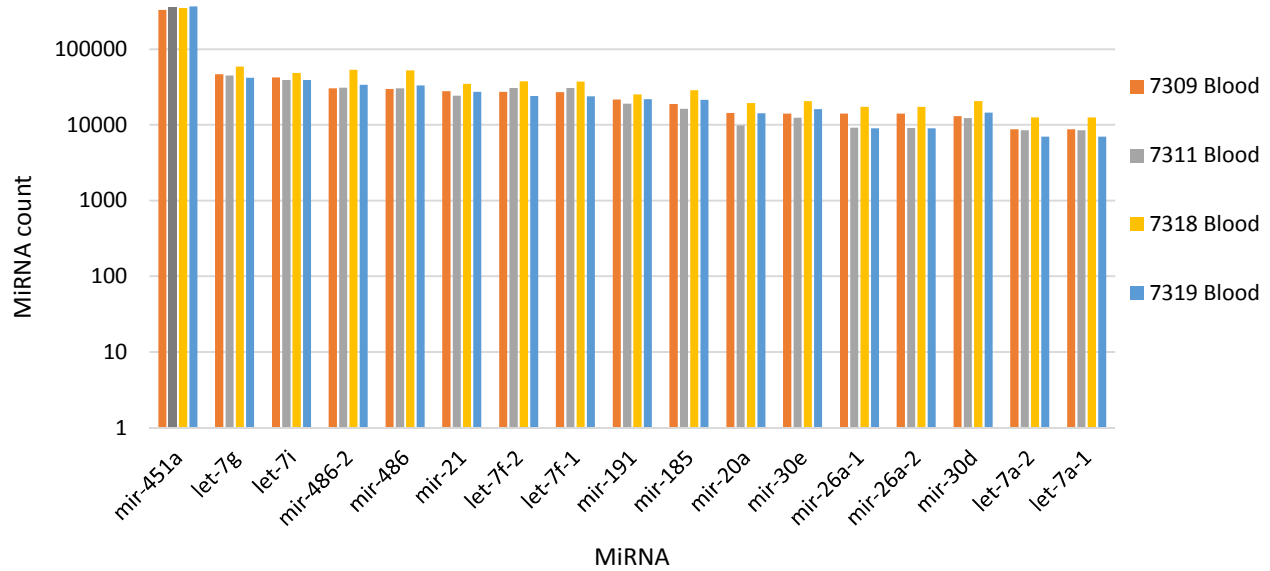
	<ul style="list-style-type: none"> • miR-30a 	<ul style="list-style-type: none"> • miR-223
Vaginal fluid	<ul style="list-style-type: none"> • miR-21 • miR-203b • miR-451a • miR-205 • let-7b • miR-99a • let-7g • let-7i 	<ul style="list-style-type: none"> • miR-203a • miR-148a • miR-27b • miR-223 • miR-22 • miR-378a • miR-200c • miR-221 • miR-101-2 • miR-101-1 • miR-200a • miR-200b
Semen	<ul style="list-style-type: none"> • miR-21 • miR-10b • miR-99a • let-7b • miR-30a • let-7g • let-7f-1 • let-7a-1 • let-7a-3 • let-7a-2 • miR-26a-1 • miR-30d • miR-26a-2 • let-7c 	<ul style="list-style-type: none"> • miR-148a • miR-200b • miR-200a • miR-375 • miR-200c • miR-151a • miR-363 • miR-148b • miR-101-2 • miR-101-1 • miR-22 • miR-27b • miR-141 • miR-28 • miR-320a
Saliva	<ul style="list-style-type: none"> • miR-21 • miR-1246 • miR-451a • miR-26a-2 • miR-26a-1 • miR-205 • miR-191 • let-7g 	<ul style="list-style-type: none"> • miR-148a • miR-223 • miR-203a • miR-143 • miR-27b • miR-101-2 • miR-101-1 • miR-140

-
- miR-30e
 - **miR-26b**
 - let-7i
 - let-7f-2
 - let-7f-1
 - miR-30d
 - **miR-340**
 - **miR-23a**
 - miR-25
 - miR-378a
 - miR-22
 - **miR-374a**
 - **miR-125b-2**
 - **miR-345**
 - **let-7d**
 - **miR-125b-1**
 - **miR-181b-2**

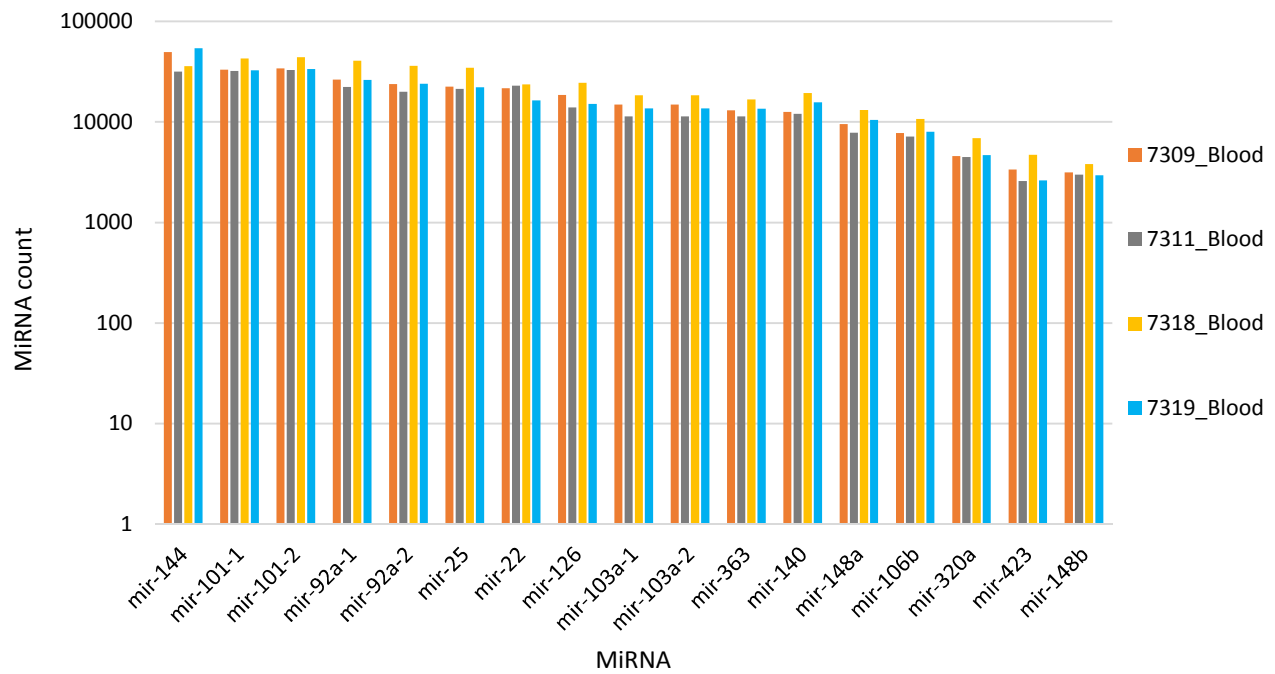
Feces

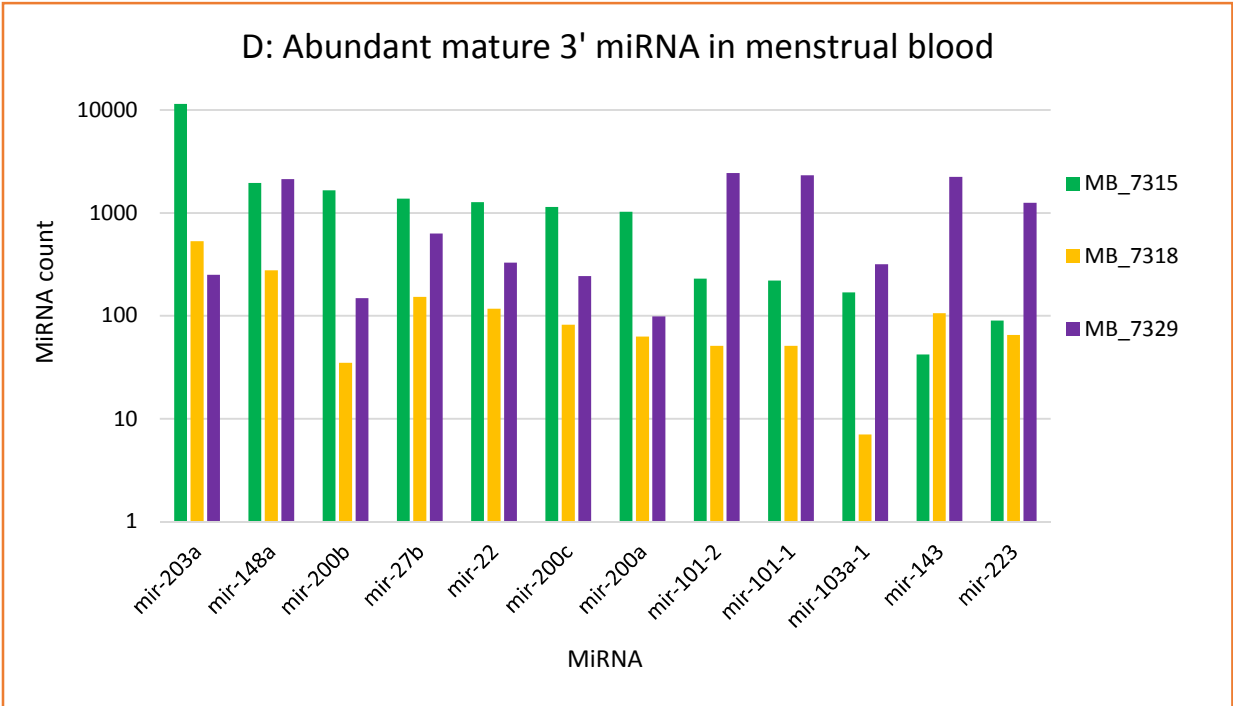
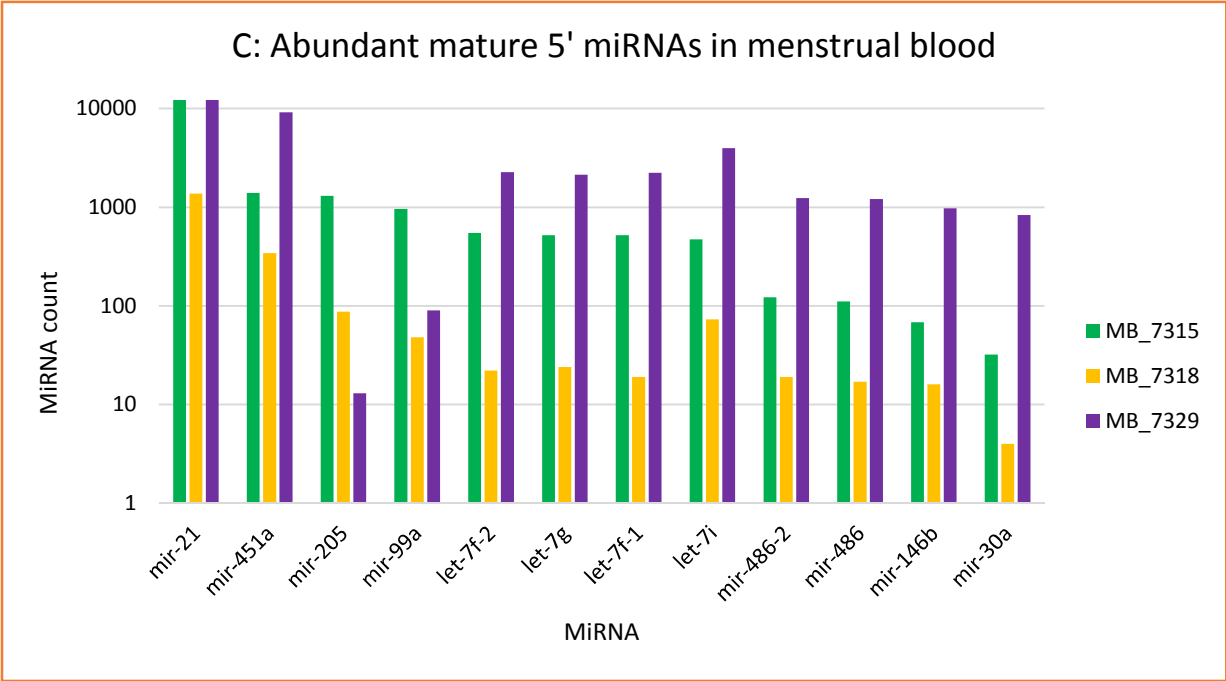
- miR-451a
 - miR-21
 - miR-486-2
 - miR-486
 - let-7i
 - let-7g
 - miR-191
 - let-7f-2
 - let-7f-1
 - miR-185
 - **miR-144**
 - miR-30d
 - miR-26a-1
 - miR-26a-2
 - miR-148a
 - miR-101-2
 - miR-144
 - miR-101-1
 - miR-25
 - miR-22
 - miR-140
 - miR-92a-1
 - miR-92a-2
 - miR-126
 - miR-103a-2
 - miR-103a-1
 - miR-203a
 - miR-363
 - **miR-106b**
-

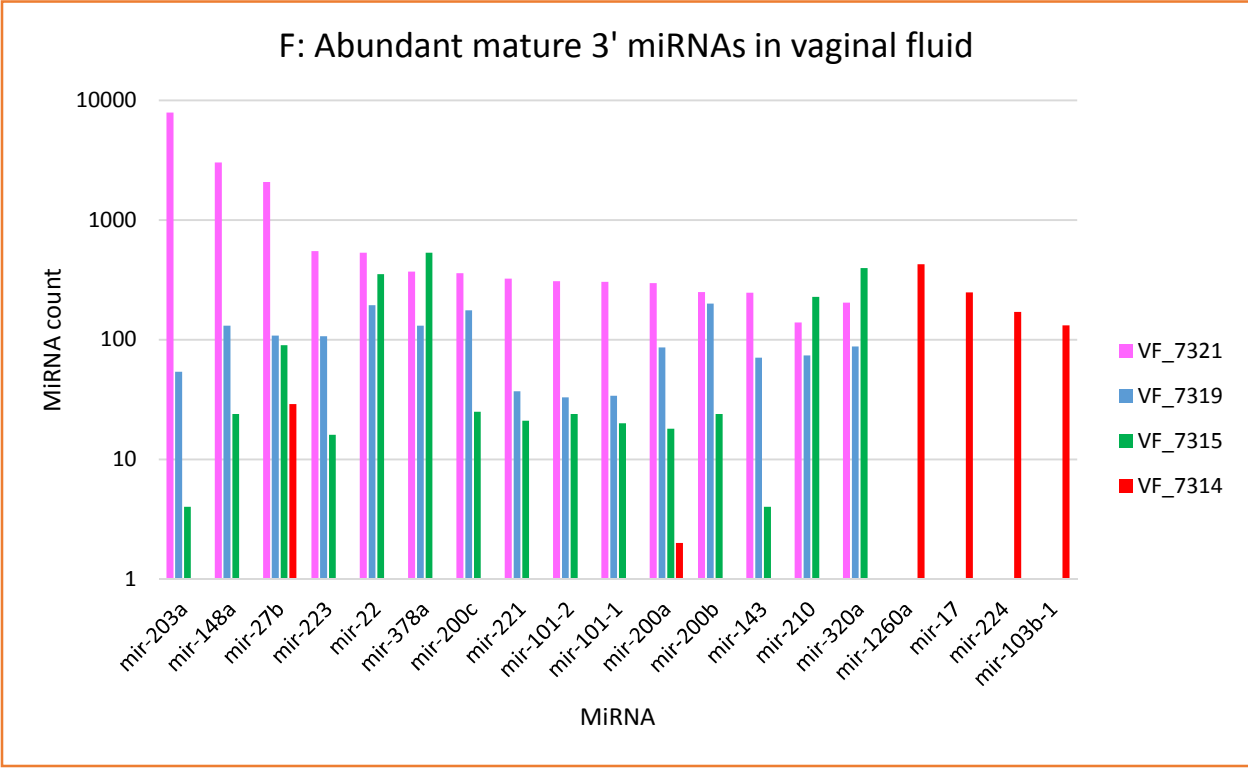
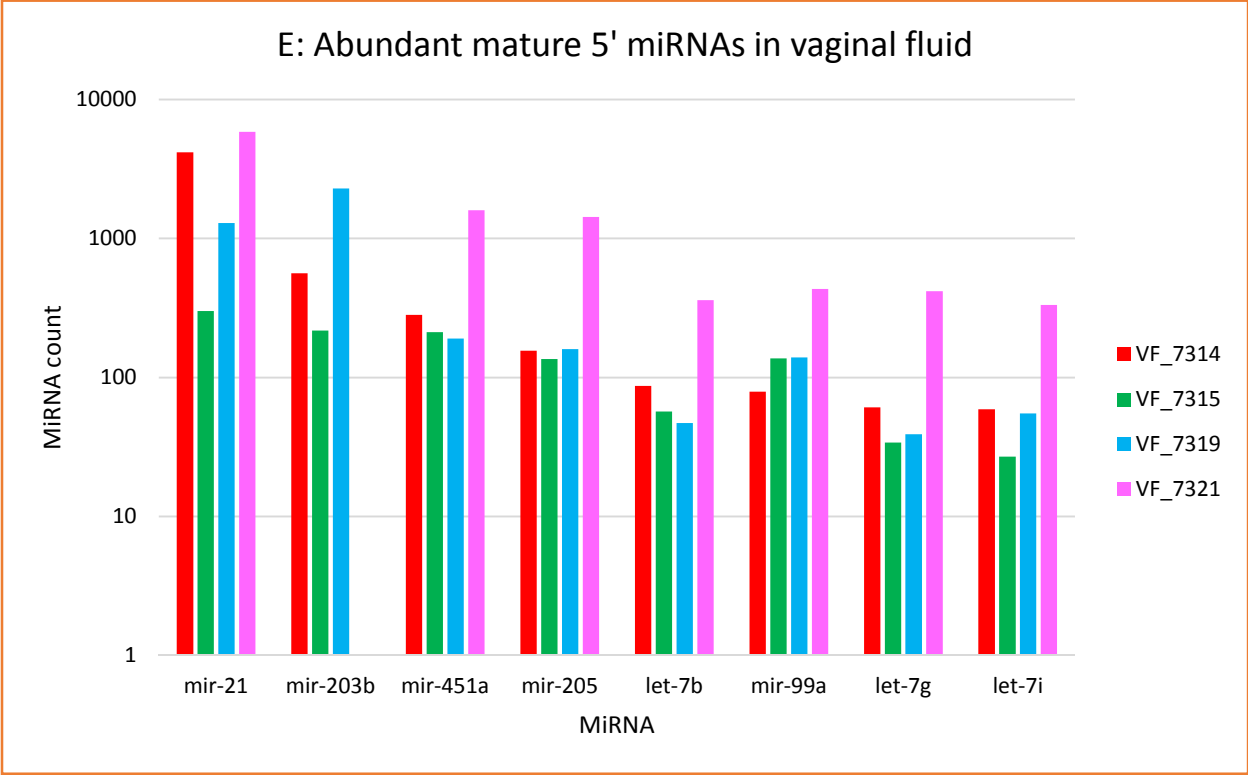
A: Abundant mature 5' miRNA in venous blood



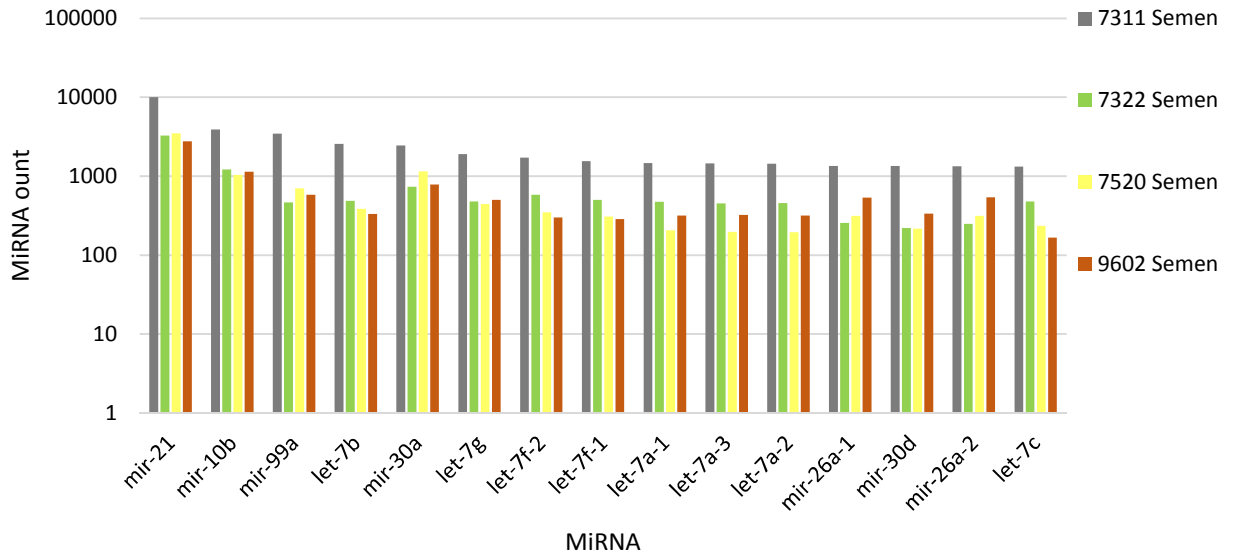
B: Abundant mature 3' miRNAs in venous blood



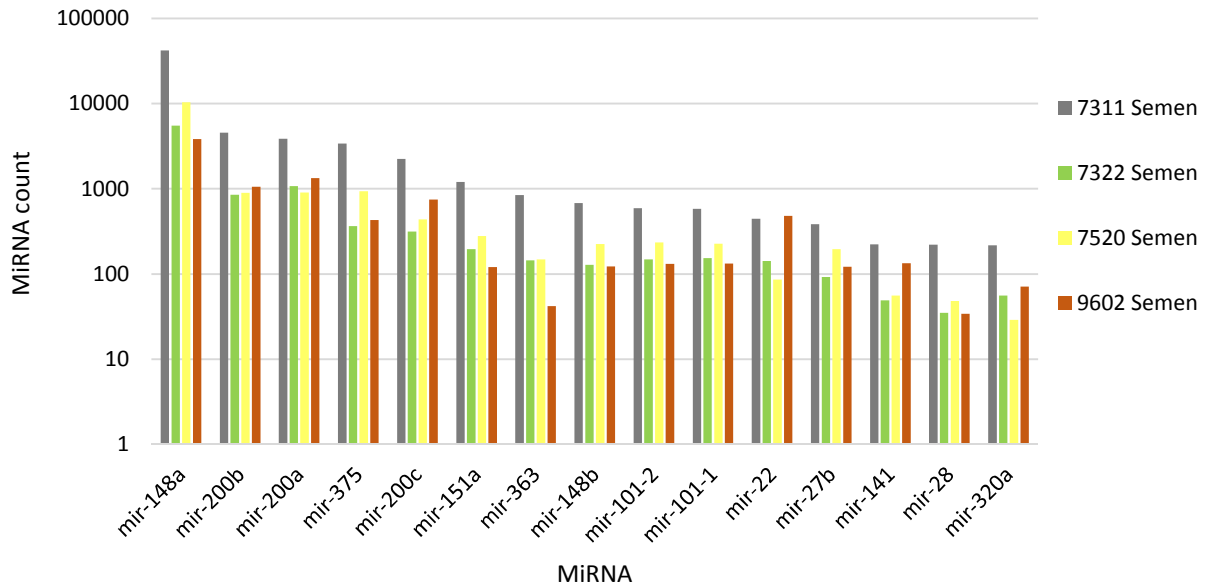


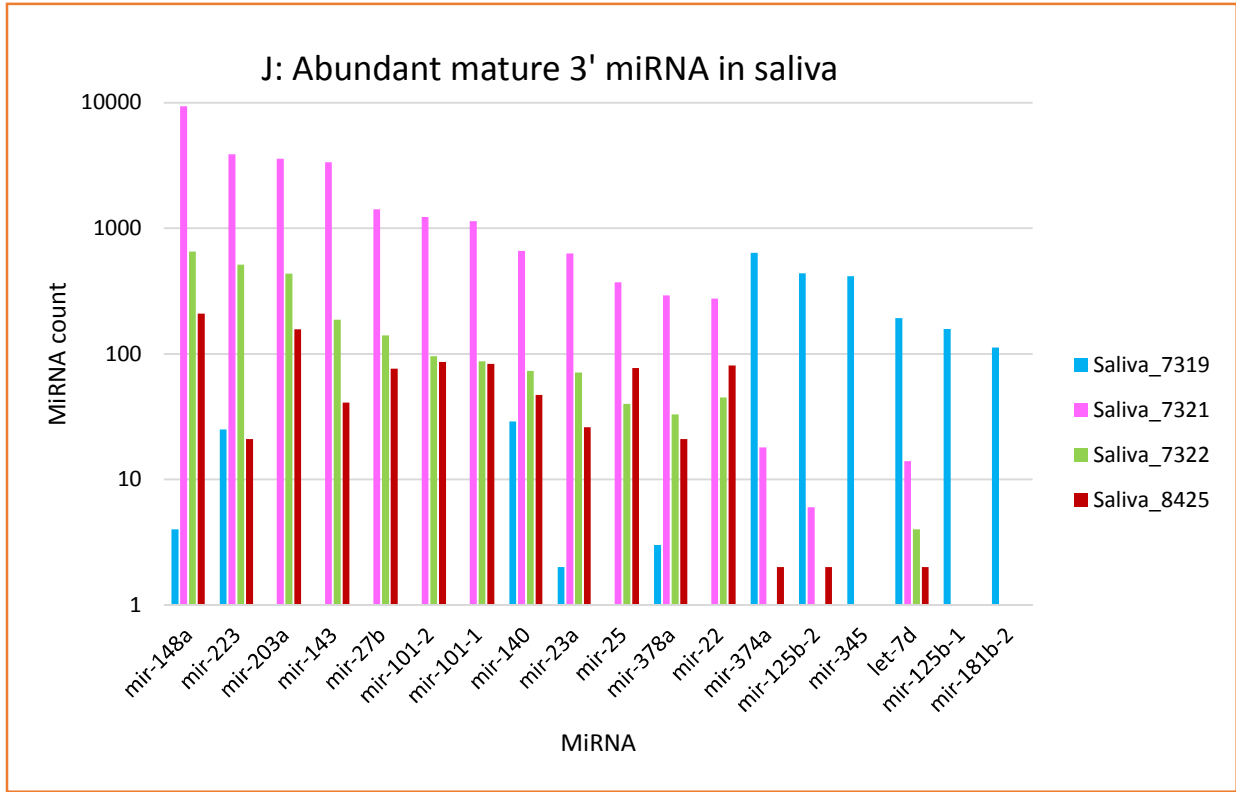
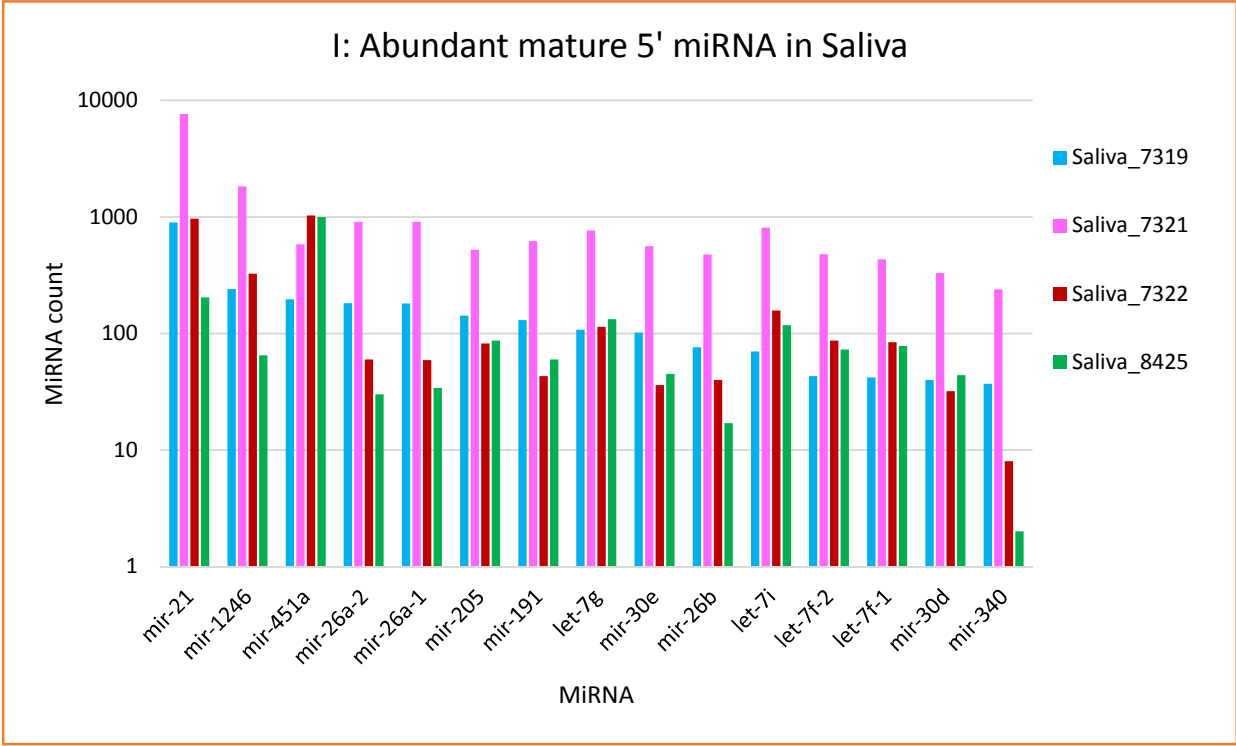


G: Abundant mature 5' miRNA in semen



H: Abundant mature 3' miRNA in semen





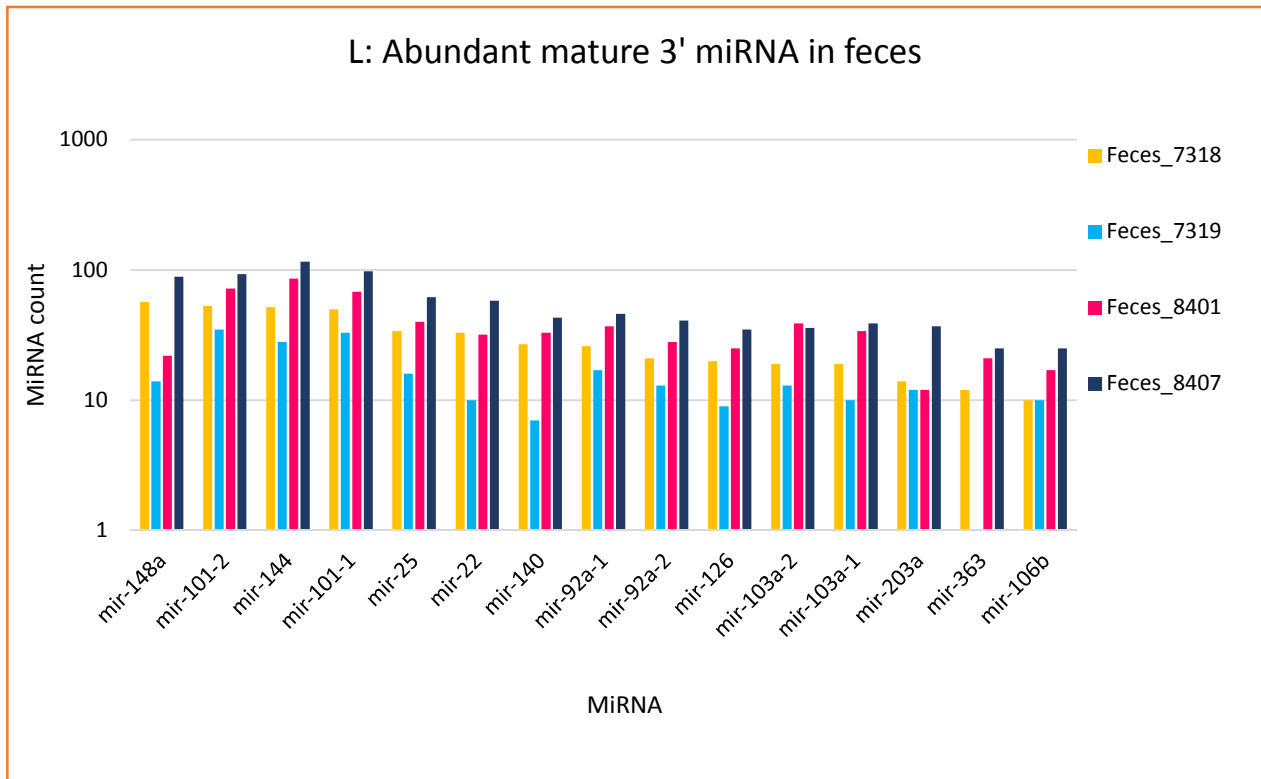
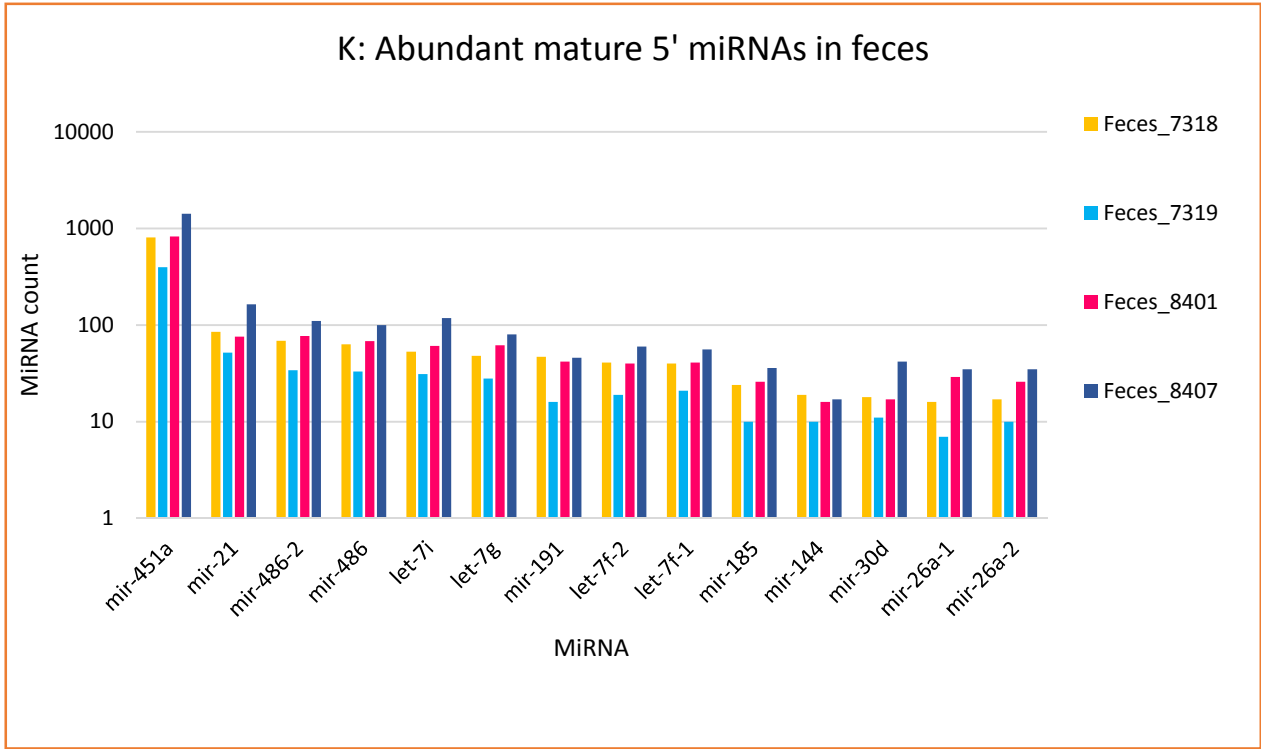


Figure 5-3 A-L: Common mature 5' and 3' miRNAs for each body fluid. MiRNA count indicates the number of times each miRNA occurred in a particular body fluid. The y-axis is listed as a logarithmic scale base 10, for better display of the small values. For each file per body fluid high count miRNAs were selected, common miRNAs were found through a UNIX command, and the data was further combined into one graph showing the pattern of how most abundant miRNAs varied per sample for each body fluid. A fixed miRNA count threshold, per file in each body fluid, was set differently, due to high variability in counts of miRNA.

5.8 Discussion:

MiRNAs are small class of non-coding RNA that regulate protein levels or inhibition of mRNA post-transcriptionally. These non-coding RNA play an important role in cellular processes such as differentiation, transformation, cell replication and regeneration [81]. Due to their broad impact a significant interest in identifying miRNAs as biomarkers for diagnoses and therapeutic monitoring of diseases such as cancer, neurodegenerative disorders, heart disease and infection has developed. Next generation sequencing (NGS) technologies are being used to identify miRNAs from biological fluids, as sequencing RNA will allow for the profiling of all extracellular miRNAs both known and unknown at one time [82]. These small non-coding RNAs are found to be extremely stable in many harsh conditions such as boiling, low or high pH, chemical or ultraviolet treatment. Studies have suggested that reason behind miRNA's exceptional stability even when treated with RNaseA is because they are not naked in nature and are encapsulated in exosomes or other microvesicles, making them highly stable and non-degradable [83].

Inspired by the idea that miRNAs can serve as biomarkers for cancer, the aim of this research was to initially sequence RNA from venous blood, menstrual blood, vaginal fluid, semen, saliva and feces of healthy male and female volunteers. In the future comparison RNA sequence data from prostate cancer patient would identify any dysregulated miRNAs that might detect and stage prostate cancer. The Illumina-HiSeq technology platform was used to sequence the RNA and the data was analyzed under the CLC genomic software. Results of this analyses showed that venous blood was rich in miRNA, its annotation percentage was the best among all the body fluids. One study reported 42% annotation of miRNAs in blood while our analyses

produced 77% annotation [84]. Most of the miRNAs observed in the published study were also observed in our analyses. To name some they are: miR-451a, miR-191-5p, miR-486-5p and miR-30d, miR-103a-3p, miR-92a-1, miR-92a-2, and miR-26a-5p. The following presents an overview of common miRNAs that were found in our analyses as well in the literature.

MiR-451a and miR-144

One study reported that miR-451a is heavily present in many body fluids, which was true for our analyses as well [84]. Except for semen, miR-451a was found in menstrual blood, vaginal fluid, saliva, and feces. MiR-144 is also found in venous blood and together with miR-451a, it is expressed under homeostatic condition, and is responsible for regulating erythroid development. In fact, another study reported that the loss of miR-144/451, causes instability in mice and increases their susceptibility to damage after being exposed to oxidant drugs. Depletion of miR-451 has also caused anemia in zebrafish embryos [85].

MiR-26a

MiR-26a was observed in venous blood, semen, and feces. This miRNA has frequently been discovered in blood and studies have shown that it is commonly down regulated in nasopharyngeal carcinoma (NPC). This miRNA reduces the expression of enhancer of zeste homolog 2 (EZH2) oncogene in NPC cells, which inhibits cell growth, and cell-cycle progression by inducing G1 phase cell-cycle arrest; whereas, overexpression of the EZH2 induces cell growth, and cell-cycle progression [86]. EZH2 is a mammalian histone methyltransferase that assists in epigenetic silencing of target genes and regulates the survival and metastasis of cancer. This gene acts as a co-activator for transcription factors such as the androgen receptor (AR). The AR plays a crucial role in prostate cancer progression and in order

to control this cancer, treatments often include inhibition of AR [87]. The correlation found between AR, EZH2 and miR-26 indicates that miR-26 can be used as a potential target for EZH2 to control prostate cancer progression and metastasis.

MiR-101-1, miR-101-2

MiR-101 was ubiquitous in all biological fluids. Studies have revealed that miR-101 also inhibits the expression of EZH2 gene in cancer cell lines. In human prostate tumors miR-101 expression decreases during cancer progression leading to an increase in expression of EZH2, which concomitantly dysregulates epigenetic pathways, causing progression and metastasis of cancer cells. The loss of expression during cancer progression indicates that miR-101 is possibly a tumor suppressor [88]. Hence, restoration of miR-101 in prostate cancer cell lines may potentially target EZH2, enabling its expression to go down, and ultimately controlling cancer progression and metastasis.

MiR-22

MiR-22 is known as metastasis-inhibitor in ovarian cancer. A negative co-relation was observed between miR-22 and the metastatic behavior of ovarian cancer cells. Bioinformatic analyses have revealed that miR-22 maybe regulating pro-metastatic genes, which may explain the inhibitory effects miR-22 has on cell migration and invasion in ovarian cancer cells [89]. Similarly another study also showed a correlation between miR-22 and cell proliferation and the tumorigenicity of hepatocellular carcinoma (HCC). In the 160 paired HCC tissue samples, miR-22 was found to be down-regulated in HCC, while histone deacetylase 4 (HDAC4) which is known to promote cancer growth was found to be up-regulated. It is possible that the loss of miR-22 contributed to the up-regulation of HDAC4, indicating that miR-22 might be targeting

HDAC4 [90]. Although, this miRNA may not have been studied in prostate cancer, but its presence in all the body fluids indicates that it is an important miRNA and its proposed effects on prostate cancer should be taken into consideration in the future.

MiR-200a, -200b, -200c family, miR-103, and miR-92a

The family of miR-200a, miR-200b, and miR-200c was observed in menstrual blood, vaginal fluid and semen. A study reported that elevated levels of the miR-200 family may correlate with serous epithelial ovarian cancer (SEOC). Where this family was found to be highly expressed in the SEOC relative to normal human ovarian epithelial cells. MiRs 103 and 92a showed no change in expression across all ovarian cell lines. Exceptional high levels of the miR-200 family members may possibly serve as biomarkers for SEOC. Whereas, miR-103 and miR-92 can be used as normalizers for RT-qPCR analyses [91]. Since the miR-200 family was observed in three of our body fluid samples, their count was not as high as compared to other miRNAs found in these body fluids. Although, the count levels of these miRNAs were low in the body fluids, it is informative to know that elevated levels of the miR-200 family may point to signs of ovarian cancer and they can be used as biomarkers for this particular kind of cancer. Similarly, miR-103 and miR-92 were observed as top miRNAs in our venous blood and feces samples but as low count miRNAs in menstrual blood, vaginal fluid and semen, therefore, they can be used as normalizers when determining the levels of miR-200 family in the above mentioned body fluids.

Mir-21

MiR-21 was another miRNA that was observed in all our body fluids. This small regulatory non-coding RNA, has been strongly conserved throughout evolution. It plays an important role in many biological functions and diseases such as development, cancer, cardiovascular diseases and inflammation. Expression of miR-21 in almost all types of cancer has been dysregulated, hence it has always been classified as an oncomiR. In particular this miRNA was found to be overexpressed in cancers such as lung, breast, stomach, prostate, colon and pancreas. Due to its consistent over-expression nature in cancers, this miRNA has also been proposed as a biomarker of malignancy in blood, sputum, cerebrospinal fluid, and feces. This miRNA has not been studied much in prostate cancer. However its involvement in tumor growth, invasion, metastasis, and cell proliferation indicates that it promotes apoptosis resistance, motility, and invasion in prostate cancer. Although heart failure in mice due to inhibition of miR-21 has proven to be beneficial, more studies need to be conducted to ascertain the therapeutic benefits of removing or inhibiting miR-21 in cancer [92, 93]. Since this miRNA has been strongly conserved evolutionarily, it has always been overexpressed in many cancers, including prostate, and was observed in all of our body fluids, this miRNA can be proposed as a potential biomarker for detecting signs of cancer in patients, in particular for prostate cancer in patients.

The presence of unique miRNAs in body fluids are as follows:

MiR-20a

MiR-20a was found unique in venous blood. The overexpression of miR-20a promotes proliferation and invasion by targeting amyloid precursor protein (APP) in human ovarian cancer cells [119]. One study screened 718 human miRNA markers in all body fluids using a microarray platform. The results indicated that several RT-qPCR confirmed miRNA markers appeared to be over-expressed in venous blood; among which four miRNAs, miR-20a, miR-106a, miR-185, and miR-144, were identified as blood markers, [120].

MiR-146b

MiR-146b was found unique in menstrual blood. A study reported that miR-146b is located on chromosome 10q24.3, and loss of this chromosome is frequently observed in gliomas. MiR-146b has been linked with the suppression of epidermal growth factor in human glioblastoma cell lines. In fact introduction of this miRNA in this cell line has been shown to decrease cell invasion, migration and phosphorylation of protein kinase B (AKT). These findings suggests that restoring miR-146b may be useful for treatment of invasive tumor [121].

MiR-203b

This miRNA was observed in the vaginal fluid. This is a skin specific miRNA that was found to be upregulated in psoriasis [122] and is involved in keratinocyte differentiation [123]. This miRNA has also been known to act as a novel tumor suppressor in esophageal cancer by down-regulating the expression of small GTPase RAN and miR-21 [123].

MiR-10b

MiR-10b was unique to semen, and one study reported that miR-10b was considered a semen marker [119]. Another study reported that miR-10b could be considered as a potential biomarker for breast cancer. Increased expression of miR-10b was observed in the lymphatic metastatic tissues, and could be directly assayed in the plasma to detect the lymph node status of breast cancer in patient [124].

MiR-26b

This miRNA was unique in saliva. Down-regulation of this miRNA has been linked to head and neck squamous cell carcinoma and oral squamous cell carcinoma [125]. Inhibition of miR-26b has also been known to affect pituitary tumor cell behavior through regulation of its direct target PTEN [126].

MiR-144-5p

MiR-144-5p was observed as unique in feces. A study reported that overexpression of this miRNA in feces suggested that it could be a potential candidate diagnostic marker for colorectal cancer detection. It was up-regulated in both feces samples and tissue samples from the CRC patients. This miRNA may be involved in cell death mechanisms, since it targets Apoptotic protease activating factor 1 (APAF1), a cytoplasmic protein that initiates apoptosis. The function of miR-144-5p or miR-144* is still unknown and further experimental analyses is needed to validate the actual function of this miRNA [127].

The annotation report for each of the body fluids

The NGS analyses revealed a wide range of miRNAs present in each of the body fluids. The annotation results revealed that venous blood was indeed rich in miRNAs. Other body fluids were not as rich in miRNAs as venous blood. A study performed by Stephen et al, reported that semen contains a heterogeneous population of spermatozoa, contributing to a varied population of small RNAs. Their research was to assess whether RNAs, such as mRNAs or small-non-coding-RNAs, delivered from mammalian spermatozoa played a functional role in embryo development. For this study, RNA sequencing was performed on Illumina's GAI sequencer, and the reads were mapped against the human reference genome, Hg19. The results revealed that human spermatozoa contains multiple classes of small RNAs, in fact it mostly consists of 65% of repeat associated small RNAs and only 7% miRNAs [94]. Our deep sequencing analyses closely correlated with this study and revealed that semen consisted of $\approx 8\%$ miRNAs. This shows that miRNAs are not abundant in semen, being present in only small amounts. Other body fluids such as menstrual blood, vaginal fluid, saliva and feces showed low annotation percentages for miRNAs. However, a slight correlation between our analyses and a study performed by Weber et al., was observed. This study demonstrated that through a global survey of miRNAs across various body fluids via RT-qPCR, 458 miRNAs were detected in saliva [10]. Our analyses detected 361 miRNAs in saliva. Although the methods to determine miRNAs in saliva were different between the two studies, the close proximity observed in the number indicates that saliva also might not be a very rich source of miRNAs.

Menstrual blood, in particular, did not annotate (4.4%) as well as a former study indicated (60%) [96]. One reason behind poor annotation could be due to low quality scores observed for our sequences. NGS consists of a series of steps that uniquely contributes to the

overall quality of the data set. Sequencing quality metric can provide information about the accuracy of each step in the sequencing process, such as library preparation, base calling, and read alignment [108]. Base calling accuracy is measured by the Phred Quality score (Q score). The Q score indicates the probability that a given base is called incorrectly by the sequencer. Q scores are initially based on Sanger sequencing accuracy, and although certain NGS techniques may vary from Sanger sequencing, the process to generate Phred quality scoring largely remains the same. In particular, *de novo* NGS platforms such as the Illumina, their methods to generate Phred Q score are similar to that of Sanger sequencing methods. Q scores are logarithmically related to base calling error probabilities. For instance, a Q score of 30 to a base is equivalent to the probability of an incorrect base call of 1 in 1000 bp, which means the probability of a correct base call is 99.9%. Similarly a Q score of 20 indicates the probability of an incorrect base call is 1 in 100 bp, meaning every 100 bp sequencing read will likely contain an error, thus the probability of a correct base call reduces to 99%. Therefore a score of Q30 is considered a benchmark for quality in NGS [108]. The Q score for menstrual blood was Q20. The quality score was determined by fastqc software [109], which is a quality control application for fastq files. In general, fastq formatted files contain a quality value for each base, which says how confident the sequencer was that the base call generated was correct. This software takes in the fastq formatted file and runs a series of tests to generate a comprehensive quality score report. It basically shows anything that might be unusual about the sequence [109]. Low quality scores can increase false-positive base calls, which could in turn result in inaccurate conclusions when analyzing sequencing data or validating experiments. A Q score of 20 was observed for other body fluids i.e. vaginal fluid, saliva, and feces, as well, hence that could be one reason for low annotation observed for these body fluids.

Another reason for low annotation could be due to genomic DNA, or bacterial contamination naturally found in menstrual blood, vaginal fluid, saliva, and feces. In general the human lower genital tract is constantly exposed to various microorganisms, which could infect the upper genital tract through migration. The normal microbial flora found in the vagina contains some gram negative bacteria that could infect the uterine wall after migrating from the vagina and thereby contaminate menstrual blood [110]. Other than bacterial contamination naturally found in these fluids, such contamination can also be encountered during sample collections. Improved purification steps and collection steps might be applied to avoid any sort of contamination in the future. For instance, it is recommended that immediately before collecting saliva, foods containing high sugar or caffeine should not be consumed, as they may lower saliva pH and increase bacterial growth. Any consumption of alcohol, nicotine, or over the counter prescription drug should be reported 12 hours in advance. The presence of any oral disease or injury should also be reported, a mouth should be rinsed at least 10 minutes before collecting saliva, to avoid diluting the samples [111]. Another effective method of collecting saliva is using the Oragene, it is a simple, painless procedure that requires the donor to spit into the collection device. After closing the device a solution is released from the cap to mix with the saliva. This solution stabilizes the DNA for long-term and prevents any bacterial contamination [112]. Many studies report collection of menstrual blood through a sterile swab, however these studies report different kits to extract RNA. For our analyses miRNeasy mini kit (Qiagen N.V., Venlo, The Netherlands) was used; however one study used the AllPrep DNA/RNA Micro kit (Qiagen, Dusseldorf, Germany). They modified the manufacturer's protocol slightly and found an increased improvement in RNA extraction and protection from RNase. While genomic DNA

from menstrual blood samples was removed using Turbo DNA-free kit (Ambion, Austin, TX, USA) [113].

In conclusion despite low annotation, miRNAs observed in menstrual blood correlated with miRNAs found by other published articles, such as miR-451a, miR-21, miR-22 [95]. In fact many other miRNAs observed in our body fluid samples also closely matched with other studies. Also, with the few selected miRNAs we were able to gain insights on miRNAs that could serve as biomarkers in prostate cancer, and some that could be used as normalizers when analyzing levels of particular miRNAs in body fluids. Overall, a wide range of miRNAs were discovered in the 6 biological fluids. To move this research further, miRNA expression levels will need to be analyzed in biological fluids from prostate cancer patients, to identify relevant biomarkers for disease validation and staging.

Chapter 6

Conclusions

MicroRNAs are a novel class of small non-coding RNA that are thoroughly being studied in cancer, including prostate cancer. Malignant cells are dependent on dysregulated miRNAs for tumor progression or metastasis. Alterations to normal cellular behavior in cancer is caused by the up-regulation of oncogenes due to loss of tumor suppressor miRNAs or down-regulation of tumor suppressor genes due to gain to oncomiRs [114]. For better treatment options for cancer, dysregulated miRNAs are now currently being considered as potential therapeutic targets. It is proposed that by restoration of lost tumor suppressors miRNAs and/or the removal of gained oncomiRs can help mediate the tumor progression and metastasis of malignant cells [115].

The majority of our laboratory work was performed with a weakly tumorigenic parental P69 cell line and its highly tumorigenic metastatic variant the M12 cell line. Using these models, a variety of dysregulated miRNAs were previously discovered. Out of a human panel of 736 miRNAs, 231 miRNA were identified as oncomiRs which exhibited a ≥ 2 fold increase in expression, and 150 miRNAs were identified as tumor suppressors which exhibited a ≤ 0.5 fold decrease in expression, in the P69 vs. M12 cell lines. A combination of a few oncomiRs and tumor suppressors were selected for single-miR confirmatory analyses, where most of the miRNAs exhibited same fold differences, but others differed. MiR-147b was consistently classified as oncomiR, while miR-299-5p was seen as a tumor suppressor in the human panel analyses, while no value was detected for the single-miR analyses. Similar dysregulated patterns for miR-299-5p was observed in laser microdissected samples from prostate cancer patient tumor biopsies as well. To either identify mRNA targets for these miRNAs or confirm their identity, a more thorough research was performed.

In this thesis, multiple miRNAs were studied in greater depth to understand their role in our prostate cancer progression model. Along with performing *in-vitro* analyses, *in silico* analyses was also performed to determine all the miRNAs present in body fluids of normal healthy individuals. MiR screen analyses can only identify a few selected miRNAs in body fluids, however NGS analyses can identify all the miRNAs. It is essential to determine which miRNAs are normally found in body fluids attained from healthy people; so that in analyzing patients' body fluids, the loss or gain of miRNAs can be diagnostic of disease.

The second chapter consisted of isolating the best miR-17-3p site(s) within the 3'UTR of ErbB2 mRNA that contributed the most in reducing its expression. ErbB2 was one of the most up-regulated proteins observed via RPMA analyses in the M12 cell line compared to its parental P69 cell line. Three potential miR-17-3p sites, 44/576, 110/576, and 451/576, were predicted and luciferase assay was performed on each of the constructs in the M12+miR-17-3p cell line. The luciferase assay confirmed a direct binding of miR-17-3p to the ErbB2 and, that deletion of second site 110/576 contributed the most to the reduced luciferase activity.

Another important miRNA studied for this thesis was miR-299-5p. This miRNA varied vastly in its nature, as for human panel analyses it expressed as a tumor suppressor, while in a couple of patient tumor biopsy samples it expressed as an oncomiR. Thus, it became important to study how this miRNA might contribute to prostate cancer progression. Through the luciferase assay, a loss in expression was observed for the OPN, thus it was confirmed that miR-299-5p can directly bind to osteopontin (OPN). OPN is present in all body fluids, and an increase in its expression has been linked to tumor invasion and metastasis in prostate [47]. To further confirm the direct binding, OPN sites were mutated and luciferase assays re-confirmed that miR-299-5p was directly targeting this mRNA. Thus, luciferase assay re-confirmed that miR-299-5p was

directly targeting this mRNA. Proliferation assay showed no change in growth upon restoration of miR-299-5p in the M12 cell line. However, migration and invasion assays for the most part revealed that miR-299-5p was acting like an oncomiR. Instead of seeing reduced invasion and migration an opposite behavior was observed. It is difficult to conclude from the conflicting data on miR-299-5p under which condition it acts as a tumor suppressor or as an oncomiR. It was felt that these results could be affected by the altered expression of only miR-299-5p, without the normal contribution of miR-299-3p, its miR-299 precursor partner. Therefore, a considerable amount of work remains to address the importance of miR-299-5p in prostate cancer.

MiR-147b appears to function as an oncomiR in prostate cancer. It has been consistently overexpressed in every type of analyses performed to identify potential miRNAs driving prostate cancer. The significance of this miRNA has not been properly studied in many cancers; however, 5 potential mRNA targets, COL4A2, ALDH5A1, NDUFA4, SDHD, and IER5, were identified for miR-147b in the literature for lung cancer [62]. The luciferase assay performed on these 5 mRNA constructs in the M12 cell line revealed a direct binding between miR-147b and the 3'UTR of 5 mRNA targets. Most of these targets are mitochondrial enzymes that are essential for normal tricarboxylic acid cycle and ETC activity [116]. The luciferase assay validates a direct binding of miR-147b to these targets, and confirms that miR-147b alters normal mitochondrial function in cancer cells, which facilitates the progression of tumor.

All miRNAs studied in this thesis are thought to promote tumor progression and enable metastasis. Prostate cancer being the second most fatal cancer for men in the U.S., requires new diagnostic methods or treatments to treat this disease. Identifying these potential miRNAs in prostate cancer has allowed us to determine a few of its potential therapeutic targets, that if blocked could help alleviate the progression of cancer. To better understand the nature of tumor progression

and driving forces behind its formation, NGS analyses was performed on normal body fluids from healthy individuals to determine the presence of abundant miRNAs. It is essential to first identify miRNAs normally present in body fluids as they will serve as biomarkers and act as positive controls against miRNAs observed in body fluids from prostate cancer patients.

A total of 6 body fluids, venous blood, menstrual blood, vaginal fluid, semen, saliva, and feces were obtained from male and female volunteers. Illumina-HiSeq technology was used to sequence the RNA obtained from these body fluids. A bioinformatic tool CLC genomic software was used to analyze the data. The analyses revealed that venous blood contained more than 70% miRNAs while feces contained only 0.2%.

Certain miRNAs that were uniquely identified for each body fluid were, miR-20a for venous blood, miR-146b for menstrual blood, miR-203b for vaginal fluid, miR-10b for semen, miR-26b for saliva, and miR-144-5p for feces. Among which miR-10b and miR-144-5p were being considered as potential biomarkers for semen and colorectal cancer respectively.

Certain miRNAs that were ubiquitous to all body fluids were: miRs-21, -22, -101-1, and -101-2. MiR-21 was being considered as a potential biomarker for cancers such as breast, prostate, and ovarian. Similarly, the family of miR-200 was observed in menstrual blood, vaginal fluid, and semen, and was being considered as biomarkers for ovarian cancer. While some miRNAs were present in all body fluids, some were missing in a couple of body fluids such as, miR-451a was missing in semen, miR-148a was missing in venous blood, miRs-92a-1 and 92a-2 were present only in venous blood and feces. MiR-451a is considered a regulator for erythroid development, while miR-92a/b are being considered potential normalizers for serum.

The research on identification of miRNAs in body fluids is currently ongoing, and to fully conclude the results, NGS analyses will have to be performed next on body fluids obtained from prostate cancer patients. This will help determine which miRNAs can be considered as biomarkers or normalizers in prostate cancer.

References:

1. Stratton, Michael R., Peter J. Campbell, and P. Andrew Futreal. "The cancer genome." *Nature* 458.7239 (2009): 719-724.
2. Hudson, Thomas J., et al. "International network of cancer genome projects." *Nature* 464.7291 (2010): 993-998.
3. Parsons, D. Williams, et al. "An integrated genomic analysis of human glioblastoma multiforme." *Science* 321.5897 (2008): 1807-1812.
4. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. *Genes Dev.*2 (2007), 1010-1024.
5. Cairns, Rob A., Isaac S. Harris, and Tak W. Mak. "Regulation of cancer cell metabolism." *Nature Reviews Cancer* 11.2 (2011): 85-95.
6. Steele, Robert, Justin L. Mott, and Ratna B. Ray. "MBP-1 upregulates miR-29b, which represses Mcl-1, collagens, and matrix metalloproteinase-2 in prostate cancer cells." *Genes & cancer* 1.4 (2010): 381-387.
7. Prostate Cancer Foundation of Australia. The Prostate- What is it?
<http://www.prostate.org.au/articleLive/pages/The-Prostate-%252d-What-is-it%3F.html>
Accessed 07/20/2014.
8. Taylor, Renea A., et al. "Human epithelial basal cells are cells of origin of prostate cancer, independent of CD133 status." *Stem Cells* 30.6 (2012): 1087-1096.
9. American Cancer Society. What is prostate cancer?
<http://www.cancer.org/cancer/prostatecancer/overviewguide/prostate-cancer-overview-what-is-prostate-cancer> Accessed 07/20/2014
10. Prostate cancer Foundation. What is prostate cancer?
http://www.pcf.org/site/c.leJRIROrEpH/b.5802045/k.6D36/What_Is_Prostate_Cancer.htm Accessed 07/20/2014
11. American Cancer Society. Key Statistics of prostate cancer.
<http://www.cancer.org/cancer/prostatecancer/overviewguide/prostate-cancer-overview-key-statistics> Accessed 07/20/2014

12. American Cancer Society. How is prostate cancer staged?
<http://www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-staging>.
Accessed 07/20/2014
13. American Cancer Society. How is prostate cancer staged?
<http://www.cancer.org/cancer/prostatecancer/moreinformation/prostatecancerearlydetection/prostate-cancer-early-detection-tests>. Accessed 07/20/2014
14. Schröder, F. H., et al. "The TNM classification of prostate cancer." *The Prostate* 21.S4 (1992): 129-138.
15. Henry, R. Y., and D. O'mahony. "Treatment of prostate cancer." *Journal of clinical Pharmacy and Therapeutics* 24.2 (1999): 93-102.
16. Xie, B. X. *et al.* "Analysis of differentially expressed genes in LNCaP prostate cancer progression model." (2011) *J. Androl.* 32, 170-182.
17. Taplin, M. E. *et al.* "Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer." (1995) *N. Engl. J. Med.* 332, 1393-1398.
18. Visakorpi, Tapio, et al. "In vivo amplification of the androgen receptor gene and progression of human prostate cancer." *Nature genetics* 9.4 (1995): 401-406.
19. Garzon, Ramiro, George A. Calin, and Carlo M. Croce. "MicroRNAs in cancer." *Annual review of medicine* 60 (2009): 167-179.
20. Poy MN, Eliasson L, Krutzfeldt J, et al. "A pancreatic islet-specific microRNA regulates insulin secretion." (2004) *Nature* 432:226–307.
21. Landthaler M, Yalcin A, Tuschl T. "The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis." (2004) *Curr. Biol.*14:2162–67
22. Carlos E. *et al.* "The role of MicroRNAs in Cancer." (2007) *Yale Journal of Biology and Medicine* 79(3-4), 131-140.
23. Lee RC, Feinbaum RL, Ambros V. "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." (1993) *Cell.* 75(5), 843–854.
24. Reinhart BJ, *et al.* "The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans." (2000). *Nature* 403(6772), 901–906.
25. Kozomara, A. & Griffiths-Jones, S. "miRBase- Release 21: integrating microRNA annotation and deep-sequencing data." (2011) *Nucleic Acids Res.* 39, D152-7.

26. Li, M., Li, J., Ding, X., He, M. & Cheng, S. Y. "microRNA and cancer." (2010) *AAPS J.* 12, 309-317.
27. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. "MicroRNA maturation: stepwise processing and subcellular localization." (2002) *EMBO J.* 21, 4663-4670.
28. Marcos, M. "MiRNAs and cancer: An epigenetics view." (2003). *Molecular Aspects of Medicine* 34, 863-874.
29. Meltzer, P.S. "Cancer genomics: Small RNAs with big impacts." (2005). *Nature* **435**, 745-746.
30. Bae, Victoria L., et al. "Tumorigenicity of SV40 T antigen immortalized human prostate epithelial cells: association with decreased epidermal growth factor receptor (EGFR) expression." *International journal of cancer* 58.5 (1994): 721-729.
31. Sobel, R. E. & Sadar, M. D. "Cell lines used in prostate cancer research: a compendium of old and new lines--part 1." (2005) *J. Urol.* 173, 342-359.
32. Zhang, X., Ladd, A., Dragoescu, E., Budd, W. T., Ware, J. L., & Zehner, Z. E. (2009). MicroRNA-17-3p is a prostate tumor suppressor in vitro and in vivo, and is decreased in high grade prostate tumors analyzed by laser capture microdissection. *Clinical & experimental metastasis*, 26(8), 965-979.
33. Ferracin M, Bassi C, Pedriali M, et al. "miR125b targets erythropoietin and its receptor and their expression correlates with metastatic potential and ERBB2/HER2 expression." (2013) *Mol Cancer* 12(1):130.
34. Seashols, S.J., "Variation and modulation of microRNAs in prostate cancer and Biological Fluids." (2013) [Dissertation]. Virginia Commonwealth University.
35. Satelli, A., and Li, S., Vimentin as a potential molecular target in cancer therapy or vimentin an overview and its potential as a molecular target for cancer therapy. *Cell Mol Life Sci* **68(18)**, 1-20 (2013).
36. Zhang, X., et al., Inhibition of vimentin or beta1 integrin reverts morphology of prostate tumor cells grown in laminin rich extracellular matrix gels and reduces tumor growth in vivo. *Mol Cancer Ther*, 2009. 8(3): p. 499-508.
37. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507-1517 (2004)

38. Scott, G. K., Goga, A., Bhaumik, D., Berger, C. E., Sullivan, C. S., & Benz, C. C. (2007). Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *Journal of Biological Chemistry*, 282(2), 1479-1486.
39. U.S. National Library of Medicine. (2014). *ErbB2*. Washington, DC: U.S. Government Printing Office.
40. Weaver, D., *A multifaceted approach identifies ErbB2 and ErbB3 proteins and microRNA-125b a key contributors to prostate cancer progression*. [Thesis]. Virginia Commonwealth University (2012).
41. Yu, D., & Hung, M. C. (2000). Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. *Oncogene*, 19(53), 6115-6121.
42. Revillion, F., Bonnetterre, J., & Peyrat, J. P. (1998). ERBB2 oncogene in human breast cancer and its clinical significance. *European Journal of Cancer*, 34(6), 791-808.
43. Yu, Dihua, and Mien-Chie Hung. "Overexpression of ErbB2 in cancer and ErbB2-targeting strategies." *Oncogene* 19.53 (2000): 6115-6121.
44. Formosa, A., et al. "MicroRNAs, miR-154, miR-299-5p, miR-376a, miR-376c, miR-377, miR-381, miR-487b, miR-485-3p, miR-495 and miR-654-3p, mapped to the 14q32. 31 locus, regulate proliferation, apoptosis, migration and invasion in metastatic prostate cancer cells." *Oncogene* (2013).
45. Chim, Stephen SC, et al. "Detection and characterization of placental microRNAs in maternal plasma." *Clinical chemistry* 54.3 (2008): 482-490.
46. Milosevic, Jadranka, et al. "Profibrotic role of miR-154 in pulmonary fibrosis." *American journal of respiratory cell and molecular biology* 47.6 (2012): 879-887.
47. Khodavirdi, Ani C., et al. "Increased expression of osteopontin contributes to the progression of prostate cancer." *Cancer research* 66.2 (2006): 883-888.
48. Tilli, Tatiana M., et al. "Osteopontin is a tumor autoantigen in prostate cancer patients." *Oncology letters* 2.1 (2011): 109-114.
49. Miller, Fred R., et al. "MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ." *Journal of the National Cancer Institute* 92.14 (2000): 1185a-1186.

50. Barnabas, Nandita, and Dalia Cohen. "Phenotypic and molecular characterization of MCF10DCIS and SUM breast cancer cell lines." *International journal of breast cancer* 2013 (2013).
51. Shevde, Lalita A., et al. "Spheroid-forming subpopulation of breast cancer cells demonstrates vasculogenic mimicry via hsa-miR-299–5p regulated de novo expression of osteopontin." *Journal of cellular and molecular medicine* 14.6b (2010): 1693-1706.
52. Östling, Päivi, et al. "Systematic analysis of microRNAs targeting the androgen receptor in prostate cancer cells." *Cancer research* 71.5 (2011): 1956-1967.
53. GeneCopoeia, manufacturer of genomics and proteomics products and research industry (2014)<<http://www.genecopoeia.com/product/search/detail.php?prt=15&cid=&key=HmiR0221&tab=qpcr>>
54. Ambros, Victor. "MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing." *Cell* 113.6 (2003): 673-676.
55. Pillai, Ramesh S. "MicroRNA function: multiple mechanisms for a tiny RNA?." *Rna* 11.12 (2005): 1753-1761.
56. ATCC. "Passage Number Effects in Cell Lines". 2010.
<<http://www.atcc.org/~media/PDFs/Technical%20Bulletins/tb07.ashx>>
57. Teferedegne, Belete, et al. "Patterns of microRNA expression in non-human primate cells correlate with neoplastic development in vitro." *PloS one* 5.12 (2010): e14416.
58. Virginia Commonwealth University, Nucleic Acids Research Facilities. Richmond Virginia.
59. Ensemble release 75, February 2014.
60. Liu, Gang, et al. "miR-147, a microRNA that is induced upon Toll-like receptor stimulation, regulates murine macrophage inflammatory responses." *Proceedings of the National Academy of Sciences* 106.37 (2009): 15819-15824.
61. Omrane, Inés, et al. "MicroRNAs 146a and 147b Biomarkers for Colorectal Tumor's Localization." *BioMed research international* 2014 (2014).
62. Bertero, Thomas, et al. "'Seed-Milarity' confers to hsa-miR-210 and hsa-miR-147b similar functional activity." *PloS one* 7.9 (2012): e44919.
63. Safran. M., GeneCards Version 3: the human gene integrator. *Database* Vol. 2010, doi:10.1093/database/baq020.

64. National Center for Biotechnology Information (2014)
<<http://www.ncbi.nlm.nih.gov/gene/1284>>
65. Gu, Xuesong, et al. "Reduced PDEF expression increases invasion and expression of mesenchymal genes in prostate cancer cells." *Cancer research* 67.9 (2007): 4219-4226.
66. Nakamura, Satoki, et al. "Transcriptional repression of Cdc25B by IER5 inhibits the proliferation of leukemic progenitor cells through NF-YB and p300 in acute myeloid leukemia." *PloS one* 6.11 (2011): e28011.
67. Wang, Liang, et al. "Polymorphisms in mitochondrial genes and prostate cancer risk." *Cancer Epidemiology Biomarkers & Prevention* 17.12 (2008): 3558-3566.
68. Dalgin, Gul S., et al. "Identification and characterization of renal cell carcinoma gene markers." *Cancer informatics* 3 (2007): 65.
69. Solaini, Giancarlo, Gianluca Sgarbi, and Alessandra Baracca. "Oxidative phosphorylation in cancer cells." *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1807.6 (2011): 534-542.
70. Moreb. J.S., et al. "Abstract 3752: The role of aldehyde dehydrogenase isoenzymes in cancer cell proliferation, migration and drug resistance." *Cancer Research*
71. Bader, A. G., et al. "Developing therapeutic microRNAs for cancer." *Gene therapy* 18.12 (2011): 1121-1126.
72. Kosaka, Nobuyoshi, et al. "microRNA as a new immune-regulatory agent in breast milk." *Silence* 1.1 (2010): 1-8.
73. Cortez, Maria Angelica, et al. "MicroRNAs in body fluids—the mix of hormones and biomarkers." *Nature reviews Clinical oncology* 8.8 (2011): 467-477.
74. Mitchell, Patrick S., et al. "Circulating microRNAs as stable blood-based markers for cancer detection." *Proceedings of the National Academy of Sciences* 105.30 (2008): 10513-10518.
75. Chen, Xi, et al. "Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases." *Cell research* 18.10 (2008): 997-1006.
76. Weber. J.A., et al. The microRNA spectrum in 12 body fluids.
77. Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics." *Nature Reviews Genetics* 10.1 (2009): 57-63.

78. Smith, T. F. and Waterman, M. S. "Identification of common molecular subsequences. *J Mol Biol*", 147(1) (1981):195--197
79. CLC workbench User Manual, CLC Genomics Workbench 6.5 (2014).
80. Lawrie, Charles H., et al. "Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma." *British journal of haematology* 141.5 (2008): 672-675. (80)
81. Etheridge, Alton, et al. "Extracellular microRNA: a new source of biomarkers." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 717.1 (2011): 85-90.
82. Burgos, Kasandra Lovette, et al. "Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing." *Rna* 19.5 (2013): 712-722.
83. Shaffer, Jonathan, Martin Schlumpberger, and Eric Lader. "miRNA profiling from blood—challenges and recommendations." (2012): 1-10.
84. Cheng, Lesley, et al. "Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood." *Journal of extracellular vesicles* 3 (2014).
85. Yu, Duonan, et al. "miR-451 protects against erythroid oxidant stress by repressing 14-3-3 ζ ." *Genes & development* 24.15 (2010): 1620-1633.
86. Lu, Juan, et al. "MiR-26a inhibits cell growth and tumorigenesis of nasopharyngeal carcinoma through repression of EZH2." *Cancer research* 71.1 (2011): 225-233.
87. Xu, Kexin, et al. "EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent." *Science* 338.6113 (2012): 1465-1469.
88. Varambally, Sooryanarayana, et al. "Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer." *science* 322.5908 (2008): 1695-1699.
89. Li, Jun, et al. "An inhibitory effect of miR-22 on cell migration and invasion in ovarian cancer." *Gynecologic oncology* 119.3 (2010): 543-548.
90. Zhang, J., et al. "microRNA-22, downregulated in hepatocellular carcinoma and correlated with prognosis, suppresses cell proliferation and tumourigenicity." *British journal of cancer* 103.8 (2010): 1215-1220.

91. Kan, Casina WS, et al. "Elevated levels of circulating microRNA-200 family members correlate with serous epithelial ovarian cancer." *BMC cancer* 12.1 (2012): 627.
92. Kumarswamy, Regalla, Ingo Volkmann, and Thomas Thum. "Regulation and function of miRNA-21 in health and disease." *RNA Biol* 8.5 (2011): 706-713.
93. Bandyopadhyay, Sanghamitra, et al. "Development of the human cancer microRNA network." *Silence* 1.1 (2010): 6.
94. Krawetz, Stephen A., et al. "A survey of small RNAs in human sperm." *Human reproduction* (2011): der329.
95. Hanson, Erin K., Helge Lubenow, and Jack Ballantyne. "Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs." *Analytical biochemistry* 387.2 (2009): 303-314.
96. Rekker, Kadri, et al. "Circulating microRNA Profile throughout the Menstrual Cycle." *PloS one* 8.11 (2013): e81166.
97. El-Hefnawy, Talal, et al. "Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics." *Clinical chemistry* 50.3 (2004): 564-573.
98. Tsui, Nancy BY, Enders KO Ng, and YM Dennis Lo. "Stability of endogenous and added RNA in blood specimens, serum, and plasma." *Clinical chemistry* 48.10 (2002): 1647-1653.
99. Valadi, Hadi, et al. "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells." *Nature cell biology* 9.6 (2007): 654-659.
100. Ratajczak, J., et al. "Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery." *Leukemia* 20.5 (2006): 847-856.
101. Taylor, Douglas D., and Cicek Gercel-Taylor. "MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer." *Gynecologic oncology* 110.1 (2008): 13-21.
102. Skog, Johan, et al. "Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers." *Nature cell biology* 10.12 (2008): 1470-1476.

103. Resnick, Kimberly E., et al. "The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform." *Gynecologic oncology* 112.1 (2009): 55-59.
104. Peltier, Heidi J., and Gary J. Latham. "Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues." *Rna* 14.5 (2008): 844-852.
105. Ng, Enders KO, et al. "Differential expression of microRNAs in plasma of colorectal cancer patients: a potential marker for colorectal cancer screening." *Gut* (2009).
106. Wong, Thian-Sze, et al. "Mature miR-184 as potential oncogenic microRNA of squamous cell carcinoma of tongue." *Clinical Cancer Research* 14.9 (2008): 2588-2592.
107. The Neb, New England BioLabs manual:
<<https://www.neb.com/faqs/2012/07/30/which-is-the-sequence-of-the-final-pcr-product?device=pdf>>
108. Quality scores for Next Generation Sequencing.
<http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf>
109. Babraham Bioinformatics <
<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>
110. Khan, Khaleque Newaz, et al. "Escherichia coli contamination of menstrual blood and effect of bacterial endotoxin on endometriosis." *Fertility and sterility* 94.7 (2010): 2860-2863.
111. Collecting salivary blood contamination
<<http://www.salimetrics.com/assets/documents/blood-cont-saliva-collection.pdf>>
112. Koni, Anna Christina, et al. "DNA yield and quality of saliva samples and suitability for large-scale epidemiological studies in children." *International journal of obesity* 35 (2011): S113-S118.
113. Li, Yan, et al. "A strategy for co-analysis of microRNAs and DNA." *Forensic Science International: Genetics* 12 (2014): 24-29.
114. Croce, Carlo M. "Causes and consequences of microRNA dysregulation in cancer." *Nature Reviews Genetics* 10.10 (2009): 704-714.

115. Hwang, H. W., and J. T. Mendell. "MicroRNAs in cell proliferation, cell death, and tumorigenesis." *British journal of cancer* 94.6 (2006): 776-780.
116. Grosso, S., et al. "MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines." *Cell death & disease* 4.3 (2013): e544.
117. Ippolito, Joseph E., and David Piwnica-Worms. "A Fluorescence-Coupled Assay for Gamma Aminobutyric Acid (GABA) Reveals Metabolic Stress-Induced Modulation of GABA Content in Neuroendocrine Cancer." *PloS one* 9.2 (2014): e88667.
118. National Center for Biotechnology Information (2014)
<<http://www.ncbi.nlm.nih.gov/gene/1284>>
119. Fan, Xingxing, et al. "miR-20a promotes proliferation and invasion by targeting APP in human ovarian cancer cells." *Acta biochimica et biophysica Sinica* 42.5 (2010): 318-324.
120. Zubakov, Dmitry, et al. "MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation." *International journal of legal medicine* 124.3 (2010): 217-226.
121. Katakowski, Mark, et al. "MiR-146b-5p suppresses EGFR expression and reduces in vitro migration and invasion of glioma." *Cancer investigation* 28.10 (2010): 1024-1030.
122. Joyce, Cailin E., et al. "Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome." *Human molecular genetics* 20.20 (2011): 4025-4040.
123. Zhang, Fang, et al. "MiR-203 suppresses tumor growth and invasion and down-regulates MiR-21 expression through repressing Ran in esophageal cancer." *Cancer letters* 342.1 (2014): 121-129.
124. Chen, Weijie, et al. "The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers." *Tumor Biology* 34.1 (2013): 455-462.
125. Shiiba, Masashi, Katsuhiko Uzawa, and Hideki Tanzawa. "MicroRNAs in head and neck squamous cell carcinoma (HNSCC) and oral squamous cell carcinoma (OSCC)." *Cancers* 2.2 (2010): 653-669.

126. Palumbo, Tiziana, et al. "Functional screen analysis reveals miR-26b and miR-128 as central regulators of pituitary somatomammotrophic tumor growth through activation of the PTEN–AKT pathway." *Oncogene* 32.13 (2012): 1651-1659.
127. Kalimutho, Murugan, et al. "Differential expression of miR-144* as a novel fecal-based diagnostic marker for colorectal cancer." *Journal of gastroenterology* 46.12 (2011): 1391-1402.

Vita

Kanika Sharma was born on April 6th, 1990, in Kolkata, India. She graduated from Deep Run High School in Glen Allen, Virginia in 2008. She received her Bachelor of Science in Bioinformatics with minor in Chemistry from Virginia Commonwealth University in Richmond, Virginia in 2012. She participated in the Howard Hughes Medical Institute summer scholars program, which lead her to participate in the International Genetically Engineered Machine (iGEM) competition in Indianapolis.