UT GSBS Dissertations and Theses (Open Access)   Graduate School of Biomedical Sciences

5-2017

NON-CODING RNAS IDENTIFY THE INTRINSIC MOLECULAR SUBTYPES OF MUSCLE-INVASIVE BLADDER CANCER

Andrea E. Ochoa

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Bioinformatics Commons, Cancer Biology Commons, Genomics Commons, and the Molecular Biology Commons

The TMC LIBRARY
Health Sciences Resource Center

NON-CODING RNAS IDENTIFY THE INTRINSIC MOLECULAR SUBTYPES OF

MUSCLE-INVASIVE BLADDER CANCER

by

Andrea Elizabeth Ochoa, B.S.

APPROVED:

_____
David J. McConkey, Ph.D.
Advisory Professor

_____
Joya Chandra, Ph.D.
On-site Advisor

_____
Varsha Gandhi, Ph.D.

_____
Chunru Lin, M.D., Ph.D.

_____
Xiaoping Su, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

NON-CODING RNAS IDENTIFY THE INTRINSIC MOLECULAR SUBTYPES OF

MUSCLE-INVASIVE BLADDER CANCER

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Andrea Elizabeth Ochoa, B.S.

Houston, Texas

May 2017

DEDICATION

I would like to dedicate this dissertation to

my husband and family

who were always there for me when I needed them the most.

To my grandfather, Ciro Ochoa,

and great-uncle, Jose D. Ramirez,

who always believed in me and my abilities.

support and encouragement. To all of my family, especially my mom, sister, dad and grandparents, who were always there to support me with their unconditional love.

I would also like to express my immense gratitude to Cancer Answers and the families that support the Andrew Sowell – Wade Huggins Professorship/Mentorship award for providing me with such wonderful support. They are an incredible group of people, and I am truly honored to have been recognized by them.

NON-CODING RNAS IDENTIFY THE INTRINSIC MOLECULAR SUBTYPES OF MUSCLE-INVASIVE BLADDER CANCER

Andrea Elizabeth Ochoa, B.S.

Advisory Professors: David J. McConkey, Ph.D. and Joya Chandra, Ph.D.

There has been a recent explosion of genomics data in muscle-invasive bladder cancer (MIBC) to better understand the underlying biology of the disease that leads to the high amount of heterogeneity that is seen clinically. These studies have identified relatively stable intrinsic molecular subtypes of MIBC that show similarities to the basal and luminal subtypes of breast cancer. However, previous studies have primarily focused on protein-coding genes or DNA mutations/alterations.

There is emerging evidence implicating non-coding RNAs (ncRNAs), both short (miRNA) and long (lncRNA), in the regulation of various biological processes involved in cancer development and progression. The molecular mechanisms of miRNAs are relatively straightforward by inhibiting their mRNA targets, but the molecular mechanisms of lncRNAs are largely unknown. The identification of miRNAs and lncRNAs that contribute to the gene expression patterns of basal and luminal subtypes of MIBC will add another layer of subtype regulation.

In this work, we sought to study the differences in miRNA and lncRNA expression across the subtypes of MIBC. We started with TCGA's cohort of 408

tumors as a discovery cohort to identify differentially expressed miRNAs and lncRNAs that were specific to the basal and luminal subtypes of MIBC. We developed our own miRNA-sequencing data set to perform validation studies, and we found that the mRNA targets of the differentially expressed miRNAs were highly reminiscent of the already known basal and luminal subtype biology. We also developed bioinformatic analyses to extract lncRNA expression data that was used for unsupervised consensus clustering. Surprisingly, unsupervised analyses of the lncRNA expression data revealed two distinct clusters that exhibited more than 90% concordance with the subtype classifications made using mRNA expression data.

Taken together, the results presented here suggest that miRNA expression profiles, or lncRNA expression profiles, could be used as an alternative strategy to identify MIBC subtype. These findings could have significant clinical implications in the development of diagnostic tools for MIBC since miRNAs and lncRNAs are both stably expressed in body fluids.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 5 (no figures)

LIST OF TABLES

CHAPTER 1

CHAPTER 3

Chapter 1: Introduction

Background & Rationale

1. <u>Introduction</u>

1.1. Bladder Cancer

The bladder is a hollow organ that is made up of four different layers, which contain different cell types. These four layers include, from innermost to outermost, the transitional epithelium (or urothelium), a layer of connective tissue, a layer of muscle, and a layer of fatty connective tissue (Figure 1) (1). The majority of bladder cancers are transitional cell carcinomas (or urothelial carcinomas) because they arise from transitional cells in the transitional epithelium (1).

Bladder cancers can develop into either non-muscle invasive bladder cancers (NMIBCs) or muscle invasive bladder cancers (MIBCs). NMIBCs are confined to the transitional epithelium and connective tissue layer, and can either grow into the hollow center of the bladder, referred to as papillary tumors (Figure 1), or can remain as a flat tumor (Figure 1). MIBCs typically appear as flat tumors that progress to invade through the other layers of the bladder, and can eventually spread outside of the bladder to nearby lymph nodes, lungs, bones or liver.
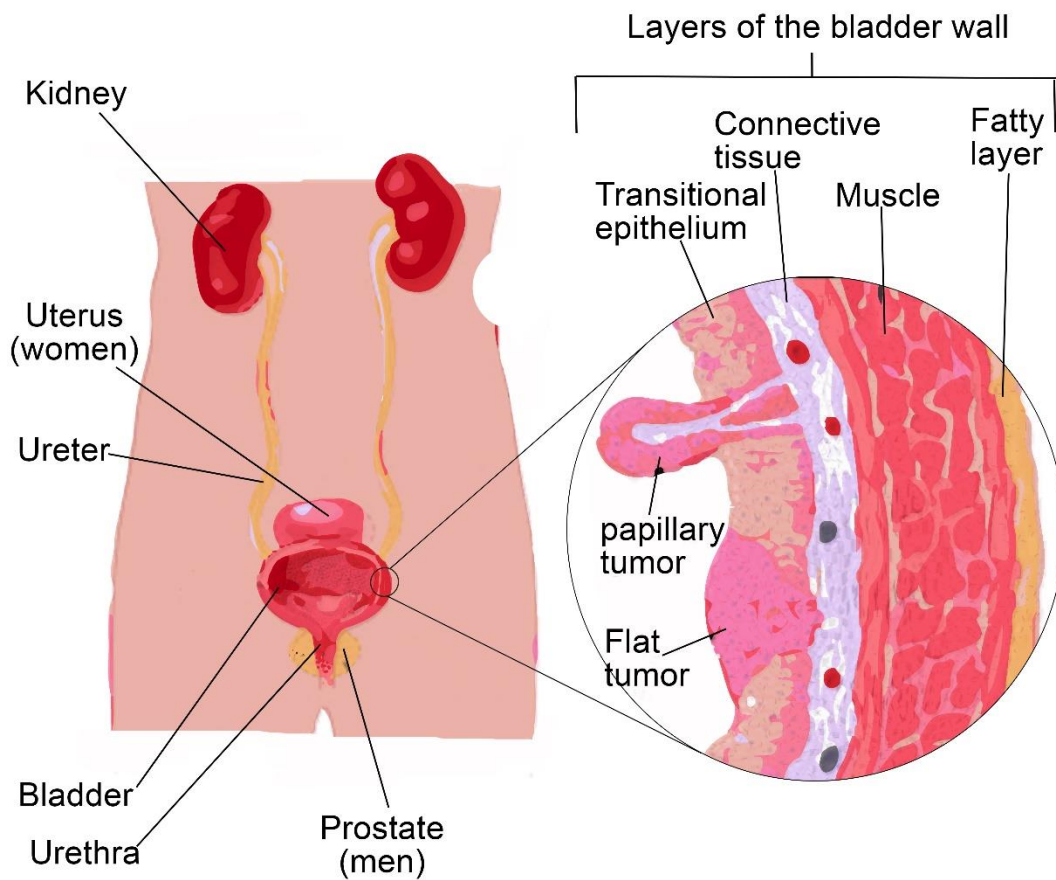
Figure 1: Illustration of the bladder wall. The four layers of the bladder wall are depicted: transitional epithelium (urothelium), connective tissue, muscle, and fatty layer. Papillary tumors are shown to grow into the hollow center of the bladder, while flat tumors do not, but have a propensity to invade into the deeper layers of the bladder wall.

1.1.1. Staging

Bladder cancers are staged based on the amount of spread into the different layers of the bladder. Non-invasive papillary carcinoma (Ta), non-invasive carcinoma in situ (Tis), and tumors that have grown into the connective tissue layer (T1) are all considered NMIBC. Tumors that have grown into the muscle layer (T2), into the layer of fatty tissue (T3), or have spread to nearby organs (T4) are all considered MIBC.

1.1.2. Statistics

Bladder cancer is the fourth most common cancer in men, occurring less frequently in women, accounting for 5% of all new cancer cases in the United States. For the year 2017, the American Cancer Society estimates 79,030 new bladder cancer cases in the United States, and 16,870 bladder cancer related deaths.

The 5-year relative survival outcomes for patients with bladder cancer vary from about 90% to 15% depending on the stage of the cancer (1). Approximately 70% of bladder cancer cases are NMIBC, with roughly 25% of those progressing to develop into MIBC. Patients with NMIBC have very good survival outcomes, with a 5-year relative survival rate around 90%, but require long-term clinical management of the disease as they are prone to recurrence (2). On the other hand, patients with MIBC exhibit a more heterogeneous spread with 5-year relative survival rates ranging from 63% to 15% (3).

1.1.3. Treatment

Bladder cancers can be treated with surgery alone or combined with other treatments, such as immunotherapy or chemotherapy. Surgery alone is often performed to treat NMIBCs, however new cancers can arise later in life. MIBCs are typically treated with a combination of platinum-based chemotherapy either before (neoadjuvant) or after (adjuvant) radical cystectomy (3, 4). Unfortunately, only about 30% to 40% of patients respond to this treatment plan, and the only alternative strategy is a recently approved anti-PDL1 immune checkpoint inhibitor (5, 6). Therefore, it is important to develop new classification systems, and identify new treatable targets, to inform and improve the clinical management of the disease.

1.1.4. Disease Characterization

Several studies have shown that various gene signatures can predict tumor stage, metastasis and progression, however the prognostication of bladder cancer is largely based on pathological criteria (7). Therefore, efforts are currently underway to integrate the pathologic evaluation with molecular features to get a more comprehensive description of a particular tumor (8). This has led to several research groups focusing on molecular data to identify tumor subtypes within MIBC in order to quickly identify the patients who are most at risk, or to identify those who would most likely respond to the available treatment options.

1.2. Molecular Subtyping

In 2000, Perou and colleagues realized that breast tumors exhibited different molecular properties affecting their responsiveness to treatments (9, 10). In order to understand the diversity of breast cancers, they characterized the gene expression patterns from 42 unique breast cancers and sought to identify molecular subtypes.

They used microarrays that measured the expression of 8,102 genes to study the gene expression patterns, and performed hierarchical clustering to group the samples. They found that the molecular portraits obtained pointed towards useful biological interpretations, such as signaling pathways. They identified 5 intrinsic subtypes of breast cancer that were related to different molecular features: basal-like, ERBB2+, normal breast-like, luminal A, and luminal B. They note a distinct difference between the two luminal subtypes, which are estrogen receptor (ER) positive, and the basal-like and ERBB2+ subtypes, which are ER negative, and the importance to treat these two subtypes as distinct diseases.

This was the first study that explored the heterogeneity within a specific cancer type, and indicated that cancers of a particular organ are clinically distinct.   By identifying subtype membership, clinicians are able to identify potential treatment strategies that will work best. By using breast cancer as a model, similar approaches have been taken to identify molecular subtypes of MIBC in order to better understand the molecular heterogeneity within the disease.

1.2.1. M.D. Anderson (MDA) oneNN classifier

Using the work done to identify intrinsic subtypes of breast cancer as a model, Choi and colleagues at MDA performed whole genome messenger RNA (mRNA) expression profiling on a cohort of 73 fresh-frozen primary MIBCs. They used Illumina BeadArrays to assess the expression of over 48,000 mRNAs, performed unsupervised hierarchical clustering, and identified three molecular subtypes of MIBC. These three subtypes are referred to as: basal, p53-like, and luminal, which were validated in an independent cohort of formalin-fixed paraffin embedded MIBCs (11).

The basal subtype showed enrichment of genes associated with basal breast cancers and squamous features (11). The basal subtype was also characterized as being more aggressive, often having metastatic disease at presentation, and shorter disease-specific and overall survival rates (11). Basal MIBCs, similar to basal breast cancers, expressed high molecular weight cytokeratins (KRT5, KRT6A/B/C, KRT14) and mesenchymal markers (ZEB2, VIM, TWIST1/2). Ingenuity Pathway Analysis of upstream regulators predicted STAT3, HIF1α, and TP63 to be transcriptional regulators of basal gene expression.

The luminal subtype of MIBC expressed the epithelial marker E-cadherin and canonical luminal biomarkers FOXA1, GATA3, ERBB2 and KRT20. Luminal MIBCs were enriched for fibroblast growth factor receptor 3 (FGFR3) mRNA expression and FGFR3 activating mutations, which are common features of NMIBCs. Ingenuity Pathway Analysis of upstream regulators implicated peroxisome proliferator-activated receptor, gamma (PPARG), ER and TRIM24 in the transcriptional

regulation of luminal MIBC. Luminal MIBCs had PPARG activation, active PPAR pathway, high PPARG mRNA expression, and high expression of PPARG's direct target and coactivator, fatty acid binding protein 4 (FABP4). Overall, when compared to the basal subtype, the luminal tumors had significantly better overall and disease-specific survival.

The p53-like subtype identified was primarily characterized by expression of an active p53 signature, but also expressed luminal biomarkers. Silhouette scores were calculated, which is a measure of subtype stability, and showed that the p53-like subtype was rather unstable, with 35% of the tumors calculated to be unstable. Despite having expression of luminal biomarkers, the p53-like tumors showed significantly poor overall and disease-specific survival when compared to the luminal tumors. Subsequent *in vitro* analyses found that these MIBCs were significantly resistant to neoadjuvant cisplatin-based chemotherapy when compared to the other subtypes.

1.2.2. UNC BASE47 subtype classifier

Damrauer and colleagues at UNC also sought to identify molecular subtypes of MIBC. They compiled samples from 3 different experiments and generated a meta-dataset of 262 MIBCs. Using whole genome expression profiling, and unsupervised consensus clustering, they identified 2 intrinsic subtypes of MIBC termed basal-like and luminal (7). Similar to results in breast cancer, and from MIBC studies at MDA, the basal-like and luminal subtypes exhibited significantly different clinical outcomes, with the basal-like subtype showing poorer outcomes. As a result of their analyses,

they developed a prediction analysis of microarrays (PAM) classifier, termed bladder cancer analysis of subtypes by expression (BASE47), which consisted of 47 genes that could accurately identify basal-like and luminal MIBCs.

They studied the genes that were significantly differentially expressed between the two clusters and identified 2,393 genes. Similar to previous findings, the basal-like subtype expressed high molecular weight cytokeratins (KRT5, KRT6B, KRT14) and CD44, while the luminal subtype expressed KRT20. The basal subtype also exhibited enrichment of genes involved in cell survival and movement.

In their study, they also emphasized the similarity of the basal-like MIBC subtype to the basal-like breast cancer subtype, and luminal MIBC to luminal breast cancer. They correlated the centroid gene expression between the bladder and breast cancer subtypes, and identified positive correlation between basal-like MIBC and basal-like and normal-like breast cancers, while luminal MIBC showed positive correlation with the luminal A and luminal B breast cancer subtypes (7). They also showed that 16% of the MIBCs shared similar characteristics to the claudin-low molecular subtype of breast cancer, which is characterized by low expression of claudins and increased expression of mesenchymal markers (7). While all of these claudin-low tumors were within the basal-like subtype, they did not show differences in clinical outcomes when compared to the rest of the basal-like tumors.

1.2.3. The Cancer Genome Atlas (TCGA)

TCGA also performed their own analysis to identify subsets of MIBC using RNA sequencing, and identified four expression subtypes. They used 129 chemotherapy-

naive, muscle-invasive urothelial carcinomas and performed mRNA, miRNA, and protein expression analyses (4).

Their Cluster I tumors were enriched for papillary morphology, FGFR3 mutations, FGFR3 copy number gain and elevated FGFR3 expression, suggesting that this subtype may respond to FGFR inhibitors. Cluster I and Cluster II expressed high levels of ERBB2 and ESR2 proteins, and showed similar features of Luminal A breast cancer, with expression of GATA3, FOXA1, and E-cadherin. Cluster III, basal/squamous-like, expressed high molecular weight cytokeratins (KRT5, KRT6, KRT14) and showed similarity to basal-like breast cancers and squamous cell cancers of the head, neck and lung (4). The Cluster IV tumors had low expression of luminal markers, moderate expression of basal markers, and relatively high expression of miR-99a and miR-100.

1.2.4. Lund subtype classification

Another group at Lund University also sought to develop a method to classify tumors based on molecular features. They used 308 bladder cancer cases, and identified 5 major subtypes: urobasal A (UroA), genomically unstable (GU), urobasal B (UroB), squamous cell carcinoma-like (SCCL), and infiltrated (Infil) (8). They performed hierarchical clustering with 308 tumor samples and used a successive two group split approach to identify their clusters. The first split identified two subtypes, which were treated individually to establish further divisions, resulting in the identification of 7 distinct subsets (8).

The UroA subtype was characterized by FGFR3 mutations, FGFR3 mRNA expression, and expression of an FGFR3 gene signature. A majority of these tumors were NMIBC, contributing to their overall good prognosis. The GU subtype was characterized by TP53 mutations, increased activity of late cell cycle genes, and expression of KRT20. Approximately 70% of the GU tumors were pathologically described as being high grade. The SCCL subtype expressed high molecular weight cytokeratins (KRT6A/B/C, KRT14), was associated with poor prognosis, and showed pathologic signs of squamous cell differentiation.   Similar to the UroA subtype, the UroB subtype had high frequency of FGFR3 mutations, but also expressed high molecular weight cytokeratins like the SCCL subtype. It was suggested that the UroB tumors may be a progressed state of UroA tumors. The Infil subtype had a strong immunologic signal, suggesting the presence of myofibroblasts, but the subtype showed quite a bit of heterogeneity because it had protein expression signatures similar to GU, UroB, and SCCL tumors.

Overall, they confirmed that the identification of molecular subtype could predict overall prognosis, with UroA showing good prognosis, GU and Infil showing intermediate prognosis, and UroB and SCCL showing the worst overall survival outcomes.

1.2.5. Summary of subtype classifications

Four different studies identifying molecular subtypes of urothelial carcinomas are highlighted here. Each group used different platforms to acquire whole genome expression data, different patient cohorts, and different methods to analyze the data.

Despite these differences, there is a remarkable amount of similarity between all of the classifiers.

Using TCGA's latest RNA-sequencing MIBC cohort (n=408) and subtype calls on this cohort from the above studies, we see that the data generally fall into two subtypes, basal and luminal (Figure 2). The MDA identified basal tumors overlap substantially with the SCCL tumors identified by Lund, and encompass all of TCGA's Cluster III and most of the Cluster IV tumors. The basal-like tumors identified by BASE47 encompass these subtypes, but also includes approximately half of the MDA p53-like tumors, and half of TCGA's Cluster II tumors. The MDA identified luminal tumors overlap with the UroA and GU tumors identified by Lund, and encompass TCGA's Cluster I and some of the Cluster II tumors. The BASE47 luminal subtype encompasses MDA's luminal tumors, half of MDA's p53-like tumors, Lund's GU, UroA, and a small portion of the Infil subtype. The tumors identified as p53-like by MDA encompass TCGA's cluster II and Lund's GU, UroA and Infil subtypes.

UNC

MDA

LUND

TCGA

UNC	Basal Luminal	
MDA	Basal Luminal p53-like	
Lund	GU Infil SCCL UroA UroB	
TCGA	cluster I cluster II cluster III cluster IV	
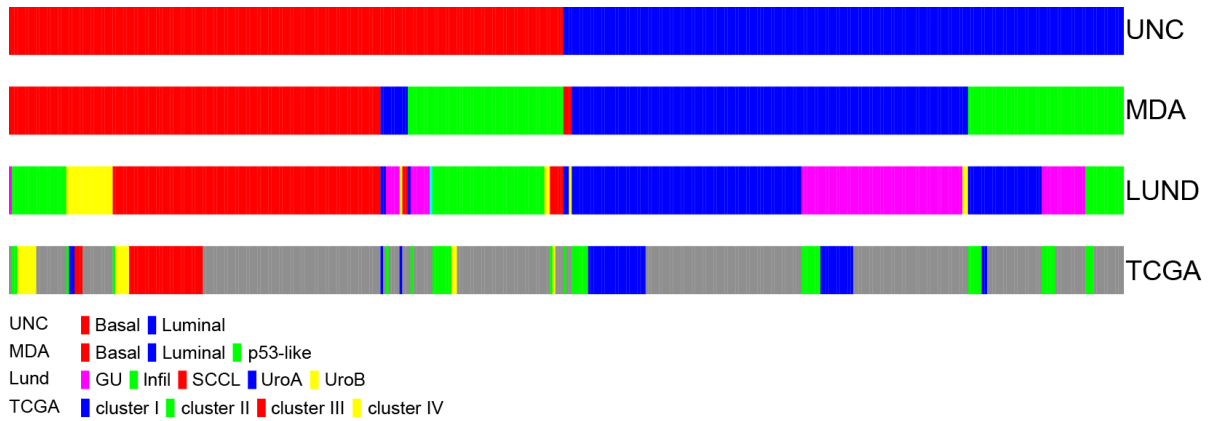
Figure 2: Summary of different subtype classifications. Using TCGA's cohort of 408 tumors we compared subtype identifications from the four different groups. UNC's 2 cluster solution: basal-like (red), luminal (blue); MDA's 3 cluster solution: basal (red), luminal (blue), p53-like (green); Lund's 5 cluster solution: GU (purple), Infil (green), SCCL (red), UroA (blue), UroB (yellow); TCGA's 4 cluster solution: cluster I (blue), cluster II (green), cluster III (red), cluster IV (yellow), missing information (gray).

1.3. Non-coding RNAs (ncRNAs)

The current understanding of the human genome is that the majority of our genome is transcribed but does not encode for protein. These portions of our genome were previously thought to be junk in the genome. However, recently this "junk" has become of major interest with the hypothesis that the non-coding regions of our genome are important and provide a hidden layer of signals that are controlling various levels of gene expression. There are many different classes of non-coding RNAs, including but not limited to transfer RNAs (tRNAs), ribosomal RNAs (rRNA), small nucleolar RNAs (snoRNAs), long non-coding RNAs (lncRNAs), and micro-RNAs (miRNAs).

Non-coding RNAs have become of prime interest in studying the heterogeneity of disease, because of their usefulness in making clinical predictions of outcome. While whole genome expression profiling data based on mRNA expression has become particularly useful to identify molecular tumor subtype, which has the potential to predict prognosis and response to treatment, this has not been easily translated into routine clinical practice worldwide mostly due to economic and technical reasons. Therefore, ncRNAs are primed to be used as potential biomarkers in minimally invasive clinical tests, as they have been shown to predict the presence of disease with high sensitivity and specificity. In this work, two classes of ncRNAs are explored, lncRNAs and miRNAs, because of their aberrant expression in several cancer types and their accessibility and stable expression in bodily fluids (12).

1.3.1. Micro-RNAs (miRNAs)

There are currently 1,881 annotated human miRNAs, according to miRbase version 21. MiRNAs are estimated to regulate approximately 30% of all human genes, with one single miRNA having the ability to affect up to 200 different mRNA targets (13). Due to the number of potential mRNA targets, miRNAs have been suggested to exert tissue-specific functions as they often target different mRNAs in different tissues. MiRNAs are small, single stranded RNAs with an approximate length of 22 nucleotides (nts). They have been described as critical for a wide range of biological processes because of their regulation of gene expression at a post-transcriptional level. In breast cancer, it has been suggested that miRNA detection in bodily fluids is far superior to mRNA profiling because of their high tissue-specificity and stability. The addition of miRNA expression to current molecular classifications may add robustness, and improve diagnosis and treatment of multiple disease types (12).

MiRNAs were first discovered in 1993 with the identification of Lin-4 in the nematode *C. elegans* (14). Several years later, in 2000, the miRNA let-7 was identified in *C. elegans* (15) and was shown to have sequence conservation in humans (16). The miRNA let-7 was detected in various human tissues and was shown to consist of 12 distinct human miRNAs within the let-7 family (15, 17). The finding that let-7 was conserved across species led to a change in the research field, and initiated research efforts focused on understanding small ncRNAs.

1.3.1.1. miRNA function

MiRNAs were found to regulate gene expression by targeting mRNAs through sequence specific targeting, resulting in mRNA degradation or translational repression. The recognition of the target mRNA is based on the complementarity of seven to eight nucleotides at the 5'-end of the miRNA (seed sequence) to the specific motif along the 3'-untranslated sequence of the target mRNA (13). Perfect, or nearly perfect, complementarity can induce degradation of the mRNA, while imperfect base pairing can result in translational inhibition (13). Translational repression occurs more frequently and is mediated by blocking the initiation step therefore affecting mRNA stability (18). MiRNAs have also been described as having decoy abilities, which is a less studied mechanism. In a decoy state, the miRNA would interfere with the function of proteins by preventing interaction between coordinating mRNAs (19).

1.3.1.2. miRNAs and cancer

The dysregulation of miRNAs in cancer was first reported in 2002, with the discovery of miR-15 and miR-16 being frequently deleted in chronic lymphocytic leukemia (CLL) (20). Since this discovery was made there have been several studies showing differential miRNA expression profiles in tumor versus normal tissues (21, 22).

MiRNAs have even been implicated in the initiation and progression of many cancers, suggesting that they play a key role in cancer biology (23). Several studies have also shown that the overexpression of a single miRNA is sufficient to initiate

tumor development (24, 25). Deregulated miRNAs can also influence tumorigenesis by decreasing or increasing the inhibition of their mRNA targets, which has led to the identification of oncogenic miRNAs and tumor suppressive miRNAs (13, 26). Oncogenic miRNAs are often upregulated in tumor tissues and target tumor suppressor genes, while tumor suppressive miRNAs are frequently downregulated in cancer and target oncogenes. However, subsequent studies have shown that miRNAs are more complex and may have dual functions depending on tumor type or stage in progression. Also, since miRNAs bind to their targets with partial, rather than perfect complementarity, there are multiple challenges with identifying mRNA targets *in silico*.

   While miRNAs are thought to act on downstream signaling, they are often involved in feedback loops, causing the expression of miRNA families to be regulated by transcription factors (27, 28). MiRNA expression can also be regulated by epigenetic modifications, as demonstrated by miRNA expression changes after exposing cells to DNA methylation inhibitors and histone deacetylase inhibitors (29, 30). While epigenetic modifications can control miRNAs, miRNAs have also been implicated in controlling epigenetic changes.

   In cancer, miRNAs have been shown to play an important role in controlling metastasis, which is one of the primary causes of cancer-related deaths. There are multiple steps involved in the metastatic process, in which miRNAs can either promote or inhibit metastasis (31). For example, miRNA-10b has been shown to positively regulate migration and invasion, and is capable of initiating metastasis in mice by targeting repressors of metastasis (32). However, more miRNAs have been

implicated in metastasis suppression, including miR-335, miR-126, miR-206, and the miR-200 family, by targeting transcription factors involved in metastasis initiation (ZEB1/2, TWIST, VIM, FN1) (33, 34).

1.3.1.3. *miRNAs and bladder cancer*

The first indication that miRNAs were deregulated in bladder cancer occurred in 2007, with the identification of 10 distinct miRNAs that were up-regulated in bladder cancers when compared to normal tissues (35). This study identified miR-223, miR-26b, miR-221, miR-103-1, miR-185, miR-23b, miR-203, miR-17-5p, miR-23a, and miR-205 to be significantly overexpressed in bladder tumors (35). Since this initial discovery, several large-scale experiments have identified differences in miRNA expression patterns across the different stages and grades of bladder cancer. Low-grade bladder cancers have been shown to have downregulation of several miRNAs, while high-grade bladder cancers are often associated with upregulation of miRNAs (36, 37). In low-grade bladder cancers, there is downregulation of miR-145, miR-143, miR-99a, and miR-100, which upregulates expression of FGFR3 (Figure 3) (36-38). In high-grade bladder cancers, there is downregulation of miR-145 to inhibit apoptosis, and upregulation of miR-21 to inhibit the p53 pathway (Figure 3) (39). Very few differences in miRNA expression were previously identified between high-grade NMIBC and MIBC.

While studies have identified differences in miRNA expression between NMIBC and MIBC, the reproducibility of these miRNA expression signatures has not been possible due to the heterogeneity of tumor specimens and research methods. The

majority of miRNA studies to date in bladder cancer consist of profiling experiments

to compare miRNA expression profiles in normal bladder versus NMIBC or MIBC

(40). While this is a good start, we have barely scratched the surface of the miRNAs

involved in bladder cancer development, progression, and the intrinsic molecular

subtypes that are present in MIBC.
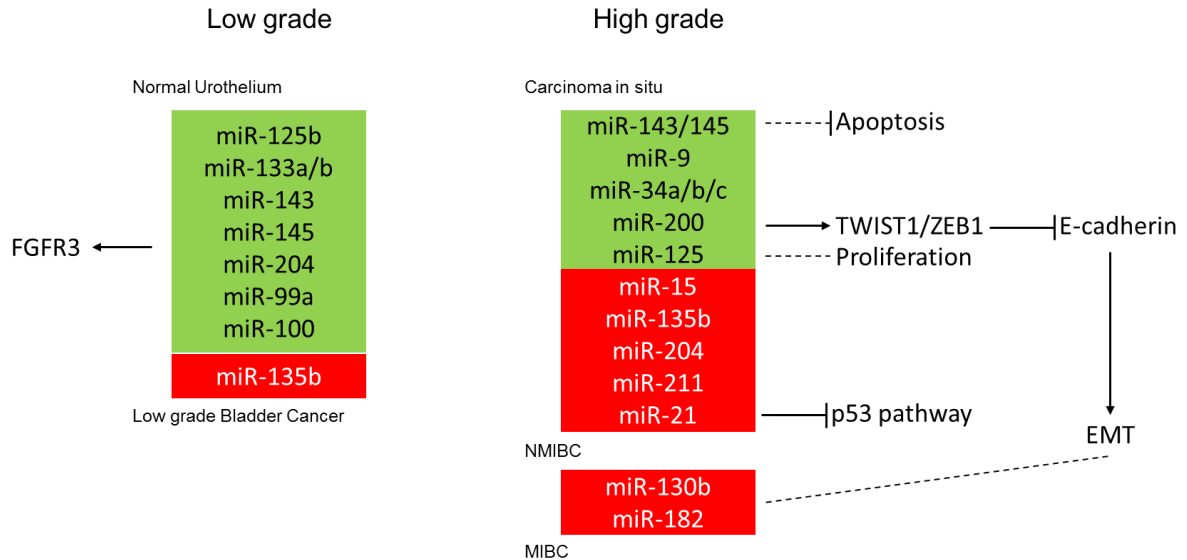
Figure 3: miRNAs identified in low-grade and high-grade bladder cancers.
(Left) As normal urothelium progresses into low grade bladder cancer there are several changes in miRNA expression. The miRNAs in the green box are downregulated, and the miRNA in the red box is upregulated in low grade bladder cancers. This miRNA expression pattern has been shown to induce FGFR3 expression. (Right) The miRNAs involved in high grade bladder cancers are distinct from the miRNAs identified in low grade tumors. The miRNA expression patterns in high grade bladder cancers are implicated in inhibition of apoptosis and the p53 pathway, and induction of EMT and proliferation.

1.3.1.4. miRNAs as biomarkers

MiRNAs have been implicated in identifying subtypes of breast cancer since miRNA expression levels were shown to predict ER, progesterone receptor (PR) and ERBB2 receptor status in breast cancer, suggesting that miRNAs might differentiate between basal and luminal subtypes (41). MiRNAs have also been associated with metastasis and epithelial-mesenchymal transition (EMT), suggesting that these miRNAs might serve as valuable biomarkers to predict metastasis and tumor recurrence.

Since cancers are commonly diagnosed through invasive techniques that involve taking a biopsy of the tumor tissue, miRNAs have been explored as potential biomarkers in bodily fluids. In cancer, the first experiment using miRNA expression profiles from the blood serum of patients was performed, and identified differences in the expression levels of three different miRNAs (42). Given the potential of miRNAs to differentiate between cancer subtypes, there is a possibility that non-invasive techniques based on miRNA expression could not only diagnose cancer, but also identify tumor subtype or patients that have a high propensity to develop recurrence. There have been a few studies that assessed the possibility of using urine from bladder cancer patients to identify miRNA biomarkers, revealing the stability of urinary miRNAs and the potential to use them in future diagnostic assays.

1.3.2. Long non-coding RNAs (lncRNAs)

The first suggestion that not all long RNA transcripts were protein-coding came with the identification of H19, a paternally imprinted maternally expressed transcript that is often deregulated in tumors (43, 44). This led to the subsequent identification of other lncRNAs, such as X inactive specific transcript (XIST) (45), and growth arrest-specific 5 (GAS5) (46).

There are currently over 21,000 identified lncRNAs, but there are likely more that have not yet been discovered since 15% of the genome remains unannotated and lncRNAs overlapping with protein-coding regions have not been analyzed yet (47). LncRNAs are arbitrarily defined as a class of non-coding RNAs that are longer than 200 nts in length, and have been implicated in a variety of biological functions by regulating the expression of genes involved in cell cycle progression, apoptosis and differentiation (47). While lncRNAs are implicated as biological signal transducers, the molecular mechanisms by which lncRNAs function still need to be elucidated. Due to the lack of knowledge surrounding the functions of the identified lncRNAs, lncRNAs are primarily characterized based on their position relative to protein-coding genes. Intergenic lncRNAs are the most commonly identified class of lncRNA, but lncRNAs can also be transcribed from exons, introns or overlapping regions of protein-coding genes (48).

LncRNAs share similar length and processing to that of protein-coding genes. While lncRNA sequences are not well conserved across species, they do exhibit tissue-specific expression but are expressed at a lower level when compared to protein-coding genes (49).

1.3.2.1. LncRNA mechanisms

While only a handful of lncRNAs have been well characterized to date, they have been shown to control every level of gene expression. They have been implicated in transcriptional gene silencing through chromatin structure, and posttranscriptional gene regulation by controlling protein synthesis and RNA transport (49, 50). In addition, many lncRNAs are thought to reside in the nucleus where they can actively regulate gene expression through chromatin remodeling complexes or potentially interact with the DNA itself (47). However, the physical association between lncRNA and chromatin modifier or gene promoter are still unclear.

LncRNAs appear to have a wide range of molecular functions, but they most likely act as a "scaffold" to assist and support the assembly of RNA-protein complexes (48). LncRNAs have also been suggested to act as guides, where they assist in the recruitment of RNA-protein complexes to target genes (48). LncRNAs may also act as a decoy by binding proteins and inhibiting the interaction between the sequestered protein and downstream target (48). While lncRNAs are primarily implicated in downstream signaling, they may also play a role upstream to control transcription factors. Overall, the mechanisms of lncRNAs are still being explored.

1.3.2.2. LncRNAs and cancer

Some of the first lncRNAs identified have been shown to be deregulated in human tumors when compared to normal controls, including H19 as a potential oncogene (43, 51-55) and GAS5 as a potential tumor suppressor (56, 57). The idea that lncRNAs exhibit cancer-specific expression was further strengthened by the

discovery of the prostate cancer antigen 3 (PCA3) lncRNA (58). It was found to be specifically overexpressed in malignant prostate tissue, and was further developed into a diagnostic tool (59). Subsequently, lncRNAs have been described in major pathways involved in cancer formation and progression.

To date, several studies have focused on the differential expression of lncRNAs in normal versus cancerous tissue, and have identified several lncRNA biomarkers associated with grade, lymph node metastasis and subtype classification (60). Overall, alterations in lncRNA expression have been associated with tumor formation, progression and metastasis by regulating gene expression involved in associated pathways.

1.3.2.3.   LncRNAs and bladder cancer

There have been 6 lncRNAs that have been implicated in bladder cancer, which are summarized in Table 1. The lncRNA H19 has been found to promote metastasis by inhibiting E-cadherin, which results in the indirect activation of pathways that promote EMT (53). H19 was also shown to increase bladder cancer growth by directly inducing the MYC transcription factor (61).

The metastasis associated lung adenocarcinoma transcript (MALAT1) lncRNA was shown to be significantly overexpressed in bladder cancers. Overexpression led to increased cell proliferation and EMT activation (62). Taurine upregulated gene 1 (TUG1) was also identified as overexpressed in bladder cancers, and was significantly associated with high-grade tumors (63). When MALAT1 and TUG1

interact, they have been shown to regulate genes involved in growth control processes.

Urothelial cancer associated 1 (UCA1) is one of the most well studied lncRNAs involved in bladder cancer. It is considered to be an oncogene since it is overexpressed in bladder cancers. Overexpression of UCA1 was shown to enhance key signaling cascades involved in cell cycle, carcinogenesis and invasion (64-66). There are three identified isoforms of UCA1, one of which was associated with cisplatin resistance and promoting tumorigenicity (67).

The maternally expressed 3 (MEG3) lncRNA has also been identified in bladder cancers, where it is considered to act as a tumor suppressor by activating p53 via MDM2 inhibition (68). MEG3 expression levels were shown to be significantly reduced in bladder cancers compared to normal controls (69), which led to increased cellular proliferation and autophagy activation (70).

LncRNAs identified in Bladder Cancer vs. Normal Controls

LncRNA	Function	Evidence
H19	Oncogene	Targets E-cadherin, induces EMT
MALAT1	Oncogene	Induces cell proliferation, migration, EMT through WNT signaling
SNHG16	Oncogene	Positive association with chemo-resistance and aggressive disease
TUG1	Oncogene	Associated with high-grade tumors and stage
UCA1	Oncogene	Promotes cell cycle, progression, and invasion
MEG3	Tumor suppressor	Activates p53, inhibits autophagy and proliferation

Table 1: LncRNAs identified to have oncogenic or tumor suppressive roles in bladder cancer. LncRNAs previously identified in bladder cancer versus normal controls, and their associated role as an oncogene or tumor suppressor.

1.3.2.4. LncRNAs as biomarkers

The altered expression of lncRNAs in urologic cancers, and their demonstrated involvement in cancer associated cellular processes, present them as attractive non-invasive biomarker candidates. This opens up the possibility for the development of clinical diagnostic tools, and possibly novel therapeutic strategies.

LncRNAs have become of interest because of their detection in bodily fluids, and their potential use as diagnostic or prognostic biomarkers. For example, HOX transcription antisense RNA (HOTAIR) was one of the first lncRNAs identified, and it was found to be associated with invasion and metastasis in breast cancers (12). The first lncRNA that was approved as a diagnostic urinary biomarker was PCA3 expression in prostate cancer (59). In bladder cancers, UCA1 expression has been measured in urine sediments, resulting in a sensitive and specific diagnostic tool to identify bladder cancer (71).

Since lncRNAs are expressed in a highly tissue, disease, or developmental-specific manner, they could serve as highly specific diagnostic, prognostic, or predictive biomarkers. With the rising number of newly discovered lncRNAs, and the accumulating evidence elucidating their functionality, there is promise for their future use in clinical practice.

1.3.3. ncRNA summary

The conventional understanding of gene regulation in biology has centered around protein-coding genes. However, over several years, evidence has suggested that the evolution and development of processes regulating the

complexity of the genome is mainly due to the regulatory potential of the noncoding portions of the genome (47). Only approximately 1.5% of the genome is protein coding, while the non-coding regulatory elements are transcribed into non-coding RNA, simply indicating that ncRNAs could play a significant regulatory role in complex organisms. While miRNAs are highly conserved across species and involved in transcriptional and posttranscriptional gene silencing through specific base pairing with their targets, lncRNAs are defined as transcribed RNA molecules greater than 200 nt in length that are poorly conserved across species and regulate gene expression by diverse mechanisms that are not yet fully understood (72).

1.4. Rationale

Overall, miRNAs are emerging as promising biomarkers for diagnosis, prognosis and response to therapy. MiRNAs are easily accessible, affordable, and present the opportunity for the development of non-invasive tools for personalized therapy. The accurate and convenient diagnosis for urologic cancers remains clinically challenging, therefore the biggest advantage to using miRNAs is that they are released from the tumor tissue in the plasma or urine where they can be easily collected and analyzed. Understanding the biology of bladder cancer progression and metastasis, as well as identifying critical markers of survival, is very important for the long-term purpose of discovering a cure for this disease. Given the role that non-coding RNAs may play in controlling the transcriptional regulation of the intrinsic subtypes of MIBC, we initiated the present study to explore the role of non-coding RNAs in the intrinsic subtypes.

Chapter 2:

Materials and Methods

2. <u>Materials and Methods</u>

Contents of this chapter are based on Ochoa AE, Choi W, Su X, Siefker-Radtke A, Czerniak B, Dinney C, McConkey DJ. 2016. Specific micro-RNA expression patterns distinguish the basal and luminal subtypes of muscle-invasive bladder cancer. Oncotarget 7(49):80164-80174.

2.1. Methods for Chapter 3

2.1.1. Human Specimens

Informed consent was obtained from all patients who contributed tumors to the TCGA and MD Anderson bladder cancer cohorts utilized in this study. All of the genomics studies were performed in compliance with US guidelines under approved IRB laboratory protocols.

2.1.2. Validation Cohort

62 fresh frozen (FF) patient samples were previously analyzed by Illumina HTv3 BeadChip microarray (11). The messenger RNA expression data was downloaded from GEO, dataset GSE48075. The dataset was quantile normalized and log2-transformed. Subsequent differential expression was performed using the R Bioconductor package, linear models for microarray data (limma) (73).

2.1.3. Ion Torrent small RNA sequencing

The same 62 RNA samples were used to perform small RNA sequencing on the Ion Proton. First, the percentage of small RNA within the total RNA sample was quantified using Small RNA and RNA 6000 Nano bioanalyzer chips from Agilent. After calculating the percentage of small RNA in the total RNA sample, we calculated the quantity required to reach 20 nanograms (ng) of small RNA, which was used for library preparation using the Ion Total RNA Seq v2 library preparation kit. The small RNA library was constructed by hybridizing and ligating the RNA, followed by reverse transcription to yield cDNA. The cDNA was then purified, size-selected, and barcoded. The resulting cDNA library was quantified with High Sensitivity DNA bioanalyzer chips from Agilent to determine the molar concentration of each library, and to calculate the percentage of library that is barcoded small RNAs. The cDNA libraries were then diluted to the same molar concentration, pooled, and diluted to 100 picomolar (pM). The pooled sample was then templated and sequenced with the Ion Proton. The GEO accession number for the small RNA sequencing data presented in this study is GSE84525.

2.1.4. TCGA training data set

Bladder cancer RNA and miRNA-sequencing level 3 data was extracted from the TCGA data portal. TCGA normalized RNAseq (RSEM) and miRNA-seq counts (RPM – reads per million mapped miRNA) were log2-transformed, median centered and filtered based on a fold change of or greater than 2 in at least 10% of the samples.

2.1.5. Consensus Clustering

Messenger RNAs and miRNAs that passed the filtering criteria were used for hierarchical consensus clustering (CC) in the R package ConsensusClusterPlus, with 80% resampling and 1,000 iterations (74). Read counts for both mRNAs and miRNAs were used as input for differential expression analysis in the R package, edgeR (75, 76). All analyses utilized a false-discovery rate (FDR) cutoff of 5%, and a fold change cutoff of 2.

2.1.6. Subtype Prediction

Prediction analysis of microarrays (PAM) was used to identify the minimal number of mRNAs or miRNAs that could accurately predict subtype classification using TCGA's cohort and the mRNA CC (k=2) calls as a reference (77). The analysis resulted in a 593-mRNA predictor (Δ = 6.969), and a 63-miRNA predictor (Δ = 3.898), which was validated using the 62 FF cohort.

2.1.7. Identification of downstream targets

Potential downstream target mRNAs of the differentially expressed miRNAs were analyzed using Ingenuity Pathway Analysis miRNA target filter. In order to be considered as an mRNA target, the mRNA and miRNA had to be negatively correlated.

2.1.8. Survival Analyses

Clinical data for TCGA's cohort was extracted from TCGA's data portal, and survival curves were generated in GraphPad Prism 6.

2.2. Methods for Chapter 4

2.2.1. TCGA data

Raw bladder cancer RNA-seq data (BAM files) and their related clinical data were downloaded from the Cancer Genomics Hub (CGHub) and TCGA data portal. The BAM files were converted to paired-end FASTQ files for each sample using bam2fastq. The raw paired-end reads in FASTQ format were aligned to the human reference genome, GRCh37/hg19, using MOSAIK and saved as BAM files (78). Raw counts for each gene were then generated by counting the mapped reads in mRNAs and lncRNAs annotated in GENCODE15 using the HTSeq package. We used the "union" mode in HTSeq to mask the overlapped regions between mRNA and lncRNA to overcome the issue of non-strand specific RNA sequencing in TCGA's data. Raw read count data were normalized using DESeq (79). Messenger RNAs and lncRNAs were kept for subsequent analysis if their FPKM was equal to or above 1 in at least 10% of the samples.

2.2.2. Differential expression analysis

Differentially expressed mRNAs and lncRNAs were identified using DESeq in the R statistical programming environment (79). Significantly differentially expressed

mRNAs and lncRNAs were identified using an FDR cutoff of or less than 5%, and

fold change cutoff of or more than a 2-fold difference.

2.2.3. LncRNA Consensus Clustering

Consensus hierarchical clustering was performed with 500 runs and 80%

resampling using the lncRNA expression data.

Chapter 3: Results

miRNA expression profiles identify subtypes of MIBC

Chapter 3: miRNA expression profiles identify subtypes of MIBC

Contents of this chapter are based on Ochoa AE, Choi W, Su X, Siefker-Radtke A, Czerniak B, Dinney C, McConkey DJ. 2016. Specific micro-RNA expression patterns distinguish the basal and luminal subtypes of muscle-invasive bladder cancer. Oncotarget 7(49):80164-80174.

3.1. Introduction and Rationale

MIBC is a highly heterogeneous disease where approximately half of the patients respond to treatment, while the other half does not. Recently, anti-PDL1 immune checkpoint inhibitor was approved for MIBC (5, 6), providing patients with an alternative treatment option. While prognostication is based entirely on clinical and pathological criteria, several large-scale genomics projects have provided new insights into the molecular heterogeneity of MIBC. The subtypes of MIBC that have been identified have shown similarities to the intrinsic molecular subtypes of breast cancers (4, 7, 8, 11). The "basal" (or "squamous cell like") subtype (80) is enriched with squamous histopathological features, and is often associated with advanced and metastatic disease at presentation along with shorter survival outcomes (4, 8, 11). The "luminal" (or "papillary") subtype had overall better outcomes, are enriched with papillary histopathological features and activating FGFR mutations (4, 8, 11, 81). Since the basal and luminal subtypes of MIBC display distinct biological and

clinical differences, they should be treated differently. Unfortunately, there are no reliable clinical tools that can quickly subtype MIBCs, revealing a true need in the field.

One way to address the clinical need of a subtype identifier is to explore the miRNA expression profiles of basal and luminal MIBCs. MiRNAs are attractive cancer biomarkers because of their stability in body fluids and FFPE tissues (82, 83). Since miRNA expression has primarily been explored in the context of tumor tissue versus normal tissue, we wondered whether miRNAs could also be used to identify basal and luminal molecular subtypes of MIBC. In this chapter, we used TCGA's matched whole genome mRNA and miRNA expression data and generated a new miRNA sequencing data set with 62 MIBCs from our own institution to explore the miRNA expression profiles of basal and luminal cancers.

3.2. Results

3.2.1. Identification of MIBC subtypes

TCGA's cohort is currently the largest, publicly available data set of high quality whole transcriptome (mRNA and miRNA) MIBC data with 405 tumors. We started with TCGA's RNA-seq data. Using an unsupervised approach, we used hierarchical consensus clustering (CC) to determine whether previous conclusions about the molecular subtypes of bladder cancer were reproducible. CC identified a three cluster (k=3) solution that best fit the data mathematically (Figure 4A), which was consistent with our previous conclusions (11). The three CC clusters also showed high expression of the biomarkers associated with the basal, p53-like and

luminal subtypes previously identified (11). While there were similar expression

patterns, the CC subtype assignments were only 75% identical to the one-nearest

neighbor (oneNN) classifier subtype assignments (Figure 4B) (11). Most of the

discrepancies were due to class switches between the p53-like and luminal tumors,

which is consistent with previous conclusions (11). Even though a three-cluster

solution was mathematically optimal, the basal and luminal biomarker expression

was almost mutually exclusive in the infiltrated tumors (Figure 4C/D), suggesting that

a two-cluster (k=2) solution would be biologically accurate.

A.

Basal ▮ Luminal ▮ Infiltrated ▮

B.

CC Basal			**CC Basal**		
oneNN Basal	124		BASE47 Basal	145
oneNN Luminal	11		BASE47 Luminal	6
oneNN p53-like	16			
CC Luminal			**CC Luminal**		
oneNN Basal	4		BASE47 Basal	6
oneNN Luminal	109		BASE47 Luminal	132
oneNN p53-like	25			
CC Infiltrated			**CC Infiltrated**		
oneNN Basal	9		BASE47 Basal	49
oneNN Luminal	35		BASE47 Luminal	67
oneNN p53-like	72			

C.

oneNN
BASE47
mRNACC

KRT5
KRT6A
KRT6B
KRT6C
CD44
KRT1
CDH3
KRT14
KRT16

FGFR3
FOXA1
GPX2
ERBB2
ERBB3
CYP2J2
GATA3
PPARG
KRT19
KRT7
KRT8
FABP4
KRT20
CD24
KRT18
XBP1

FLNC
ACTC1
ACTG2
CNN1
DES
MYH11
PCP4
SYNM
PGM5
MFAP4

oneNN ▮ Basal ▮ Luminal ▮ p53-like
BASE47 ▮ Basal ▮ Luminal
mRNACC ▮ Basal ▮ Luminal ▮ Infiltrated

-3.00 -2.00 -1.00 0.00 1.00 2.00 3.00

D.

oneNN
BASE47
mRNACC

AHNAK2
CHST15
MT1X
EMP3
MSN
MT2A
CDK6
ALOX5AP
PRRX1
CD14
GLIPR1
PDGFC
PRKCDBP
FAP
PALLD
TUBB6

CAPN5
UPK2
SCNN1B
PPARG
TOX3
GATA3
HMGCS2
RAB15
C10orf116
SEMA5A
TRAK1
SCNN1G
TMPRSS2
VGLL1
TBX2
UPK1A
FAM59A
BHMT
SPINK1
GPD1L
RNF128
CYP2J2
GDPD3
FBP1
SLC27A2
TMEM97
PLEKHG6
CYP4B1
SLC9A2
PPFIBP2
FAM174B

Figure 4: mRNA consensus clustering 3-cluster solution. A. Unsupervised CC with TCGA's cohort (n=405) identified a 3-cluster solution to be optimal. B. Comparison of the mRNA CC subtype assignments to MDA oneNN subtype assignments (left), and to BASE47 assignments (right). C. Visualization of the relative expression of MDA subtype biomarkers: basal biomarkers (top), luminal biomarkers (middle), p53-like biomarkers (bottom). D. Visualization of the relative expression of BASE47 genes: basal mRNAs (top), luminal mRNAs (bottom).

Consistent with the idea of a k=2 solution being biologically accurate (Figure 5A), we observed excellent overlap (93%) with the subtype assignments made using an independent basal/luminal PAM classifier (7). We also compared the basal/luminal assignments made by CC to the basal/luminal assignments made by oneNN, p53-like tumors were omitted, and observed 93% overlap (Figure 5B). Visualization of the basal and luminal biomarkers confirmed excellent separation of the basal and luminal CC subtypes. However, a small fraction of the basal tumors that corresponded to some of the oneNN identified p53-like tumors, had noticeably lower expression of both basal and luminal biomarkers (Figure 5C/D).
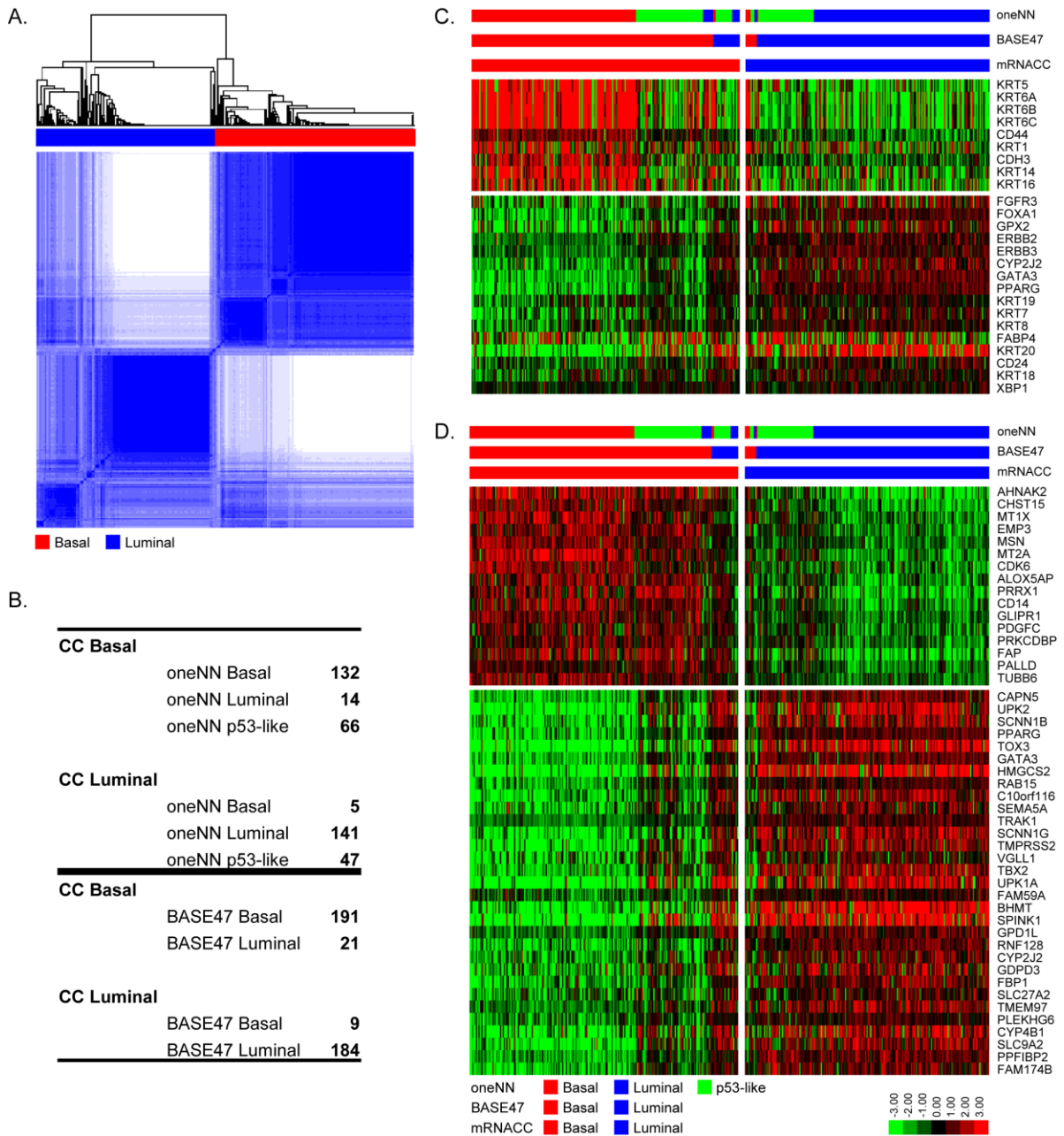
Figure 5: mRNA consensus clustering 2-cluster solution. A. Unsupervised CC with TCGA's cohort (n=405) identified a 2-cluster solution. B. Comparison of the mRNA CC subtype assignments to MDA oneNN subtype assignments (top), and to BASE47 assignments (bottom). C. Visualization of the relative expression of MDA subtype biomarkers: basal biomarkers (top), luminal biomarkers (bottom). D. Visualization of the relative expression of BASE47 genes: basal mRNAs (top), luminal mRNAs (bottom).

41

3.2.2. mRNA PAM classifier development

We used the TCGA mRNA data set and the k=2 CC subtype assignment as a
training set, to develop our own mRNA PAM classifier. For validation, we used 62
samples from the MDA FF cohort (GSE48075, Table 2). Since the whole genome
expression profiling data was generated on two different platforms, we started with
12,407 mRNAs that passed the filtering criteria in TCGA's cohort and were present
in the FF cohort to develop the PAM classifier. PAM identified 593 mRNAs that
optimally identified basal and luminal subtype classification, which contained 39 of
the 47 BASE47 genes (Figure 6A). There were a few discrepancies in the subtype
assignments made using this new PAM classifier and BASE47, which were largely
due to the presence of a small fraction of tumors that were double positive for basal
and luminal biomarkers (Figure 6).

Characteristics of the MDA validation Cohort (n=62)

	Total	Basal	Luminal	p-value
Cohort Size (n)	62	28 (45%)	34 (55%)	
Mean Age (y) ± SD	68.7 ± 9.9	69.8 ± 10.3	67.9 ± 9.6	0.686
Clinical Stage at Specimen Collection (n)				
cT2	38 (61%)	17 (61%)	21 (62%)	0.774
cT3	18 (29%)	9 (32%)	9 (26%)	
cT4	6 (10%)	2 (7%)	4 (12%)	
Positive Clinical Lymph Nodes (n)	10 (16%)	5 (18%)	5 (15%)	0.744
Positive Clinical Metastasis (n)	5 (8%)	3 (11%)	2 (6%)	0.65
Median Overall Survival (m)	46.30	18.70	82.40	0.073

Table 2: Clinicopathologic characteristics of the MDA validation cohort (n=62).
Summary of the clinical characteristics of the validation cohort used. The Mann-Whitney U test was used to compare differences in mean age between groups. The log-rank test was used to compare differences in survival between groups. Fisher's exact test was used to determine differences between groups for the remainder of categorical variables.
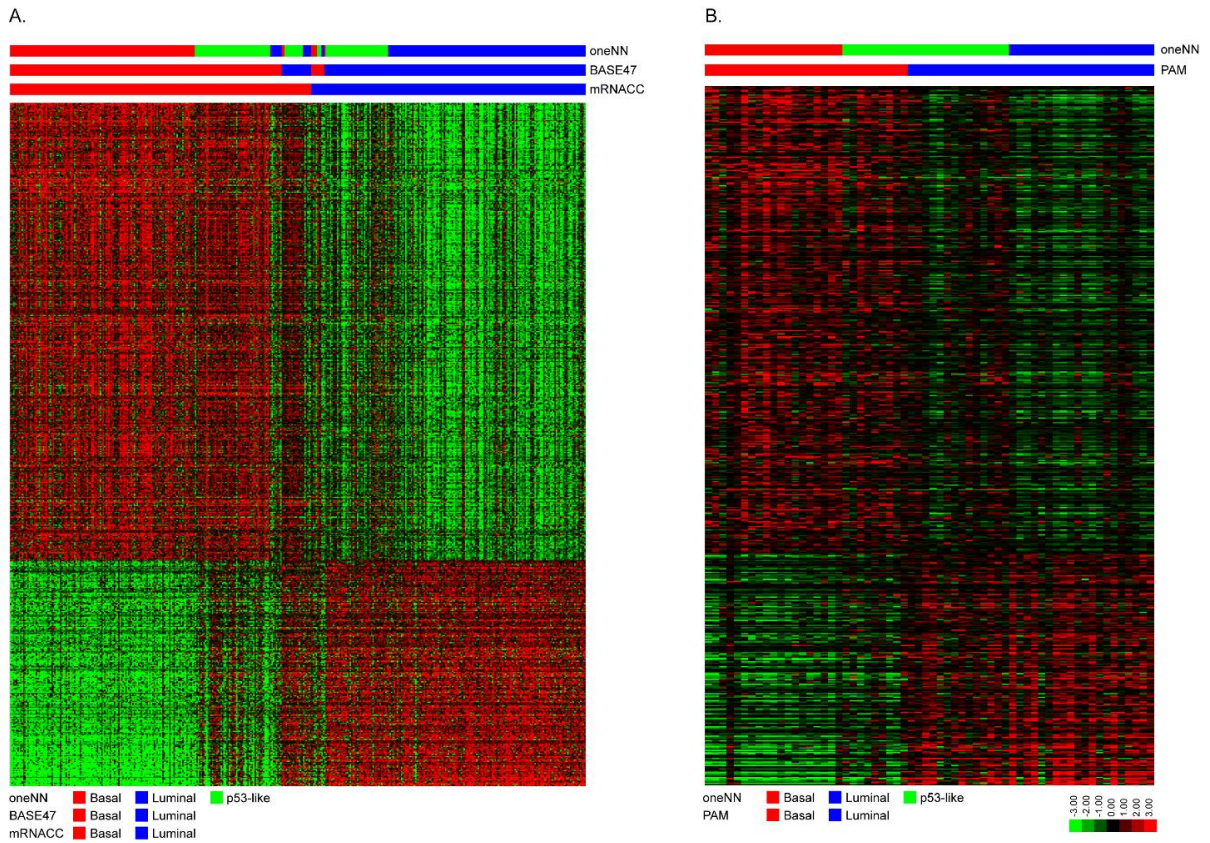
Figure 6: 593-mRNA PAM genes. A. Visualization of the 593-mRNA PAM genes in TCGA's cohort. B. Visualization of the 593-mRNA PAM genes in the FF cohort. Color bars: MDA – basal (red), luminal (blue), p53-like (green); BASE47 – basal (red), luminal (blue); mRNACC/PAM – basal (red), luminal (blue).

The 593 PAM mRNAs were then used to make subtype predictions on the 62

FF cohort (Figure 6B), resulting in 28 basal tumors and 34 luminal tumors. The 28

predicted basal tumors encompassed all the basal tumors originally identified by the

oneNN prediction, and had high expression of basal markers and basal BASE47

genes (Figure 7). The 34 predicted luminal tumors encompassed all the luminal

tumors identified by oneNN prediction and had high expression of luminal markers

and luminal BASE47 genes (Figure 7).
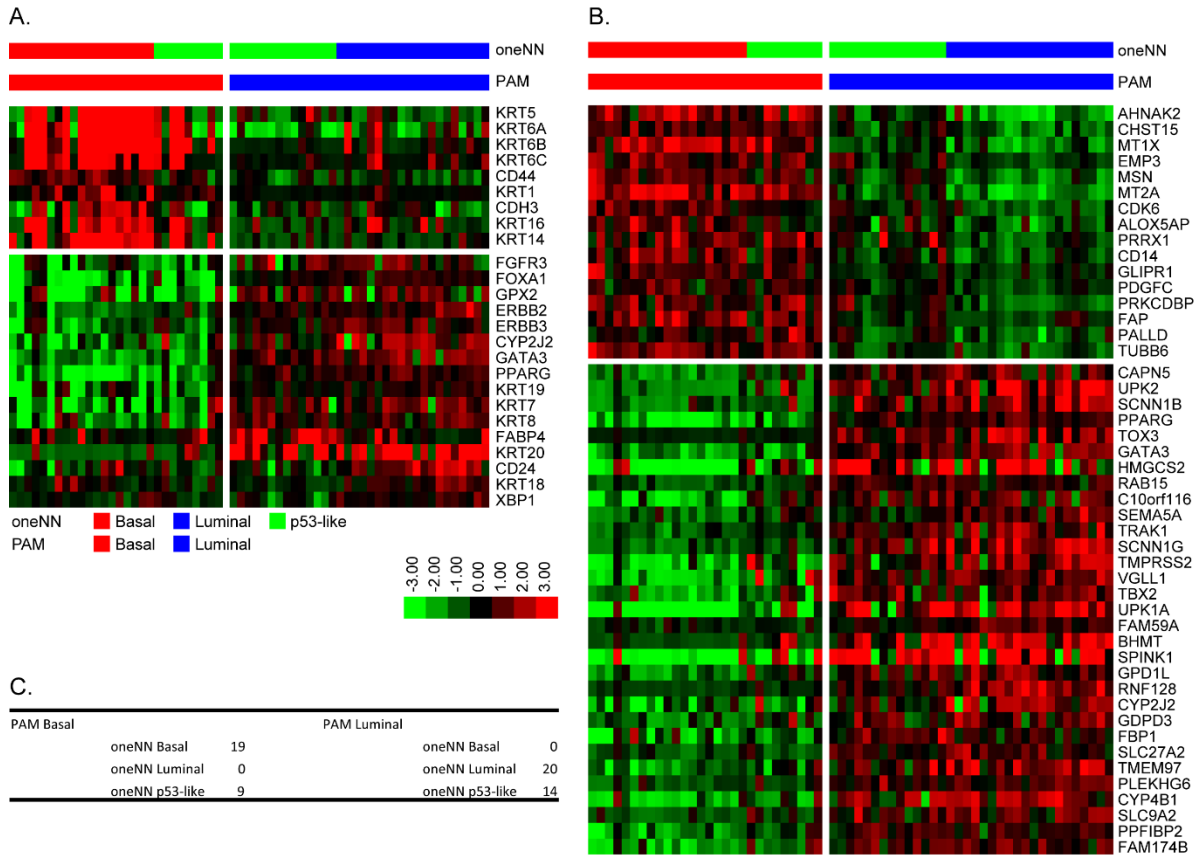
Figure 7: mRNA PAM classifier identifies basal and luminal tumors in validation cohort. A. Visualization of the relative expression of MDA basal and luminal markers as a function of PAM identified subtype classification. B. Visualization of the relative expression of BASE47 basal and luminal genes as a function of PAM identified subtype classification. C. Comparison of mRNA PAM subtype assignments and MDA oneNN assignments. Color bars: oneNN – basal (red), luminal (blue), p53-like (green); PAM – basal (red), luminal (blue).

3.2.3. Identification of miRNA based subtypes

After identifying the basal and luminal MIBCs in the two datasets, we characterized their differential miRNA expression patterns as a step toward identifying a miRNA-based signature that could distinguish the two subtypes. We used the same unbiased approach, and used TCGA's miRNA-seq data from 405 MIBCs to perform consensus hierarchical clustering (Figure 8). Using a k=2 solution (Figure 8A), similar to the mRNA analyses, we observed 77% and 73% concordance with the basal and luminal subtype assignments made by mRNA CC and BASE47, respectively (Figure 8B). However, when we examined basal and luminal mRNA biomarker expression based on the miRNA clusters, we observed a significant fraction of the mRNA PAM-defined luminal tumors clustered with the miRNA-defined basal cluster (Figure 8C/D).
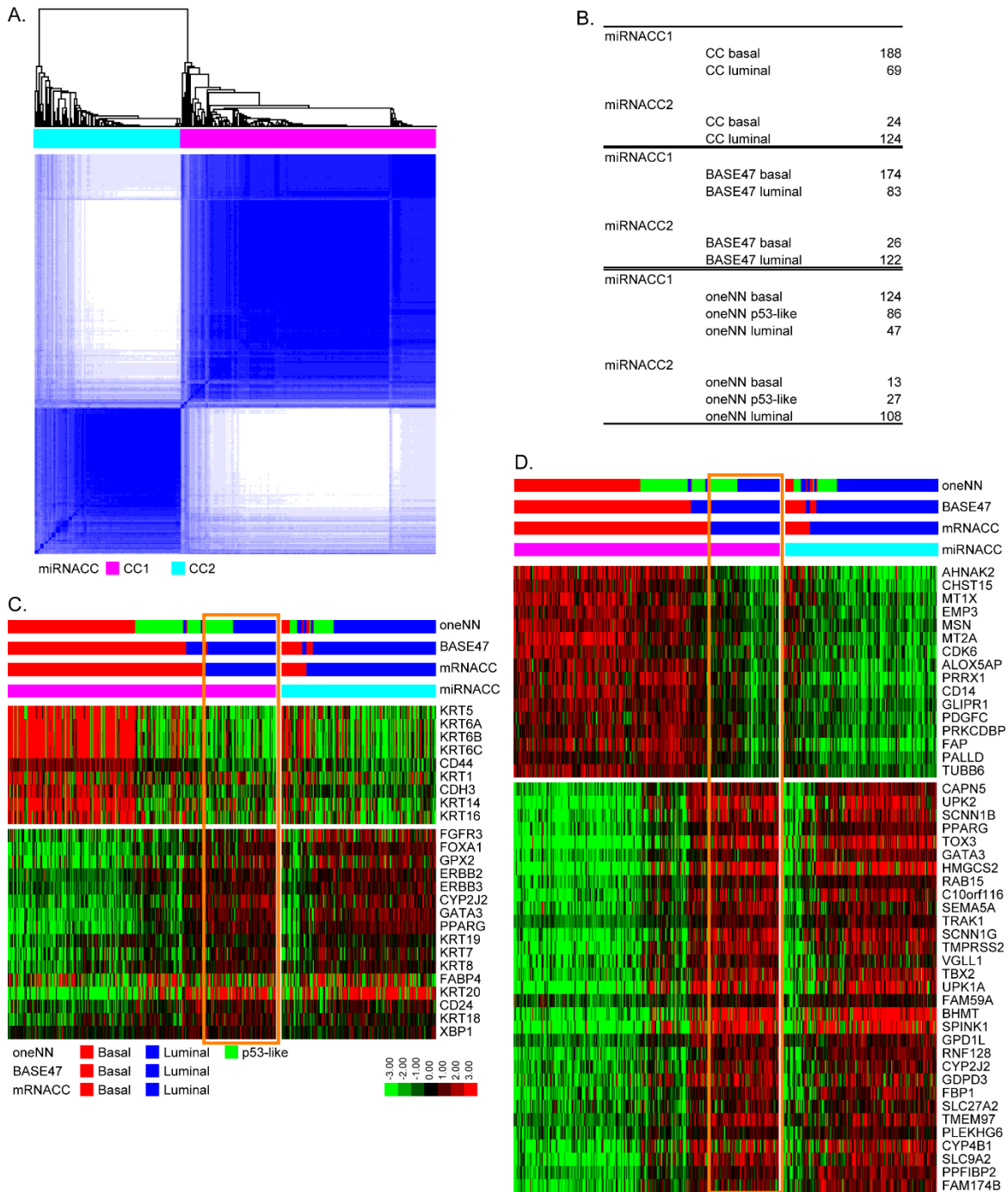
Figure 8: miRNA consensus clustering 2-cluster solution. Unsupervised CC was performed with the miRNA expression from 405 tumors in TCGA's cohort. A. MiRNA CC identified 2 distinct clusters. B. Comparison of miRNA CC subtype assignments to mRNA CC assignments (top), to BASE47 assignments (middle), and to MDA oneNN assignments (bottom). C. Visualization of the relative expression of MDA basal and luminal markers. D. Visualization of the BASE47 basal and luminal genes. Color bars: miRNA CC – basal (purple), luminal (light blue); oneNN – basal

(red), luminal (blue), p53-like (green); BASE47 – basal (red), luminal (blue); mRNACC– basal (red), luminal (blue). Note: Orange boxes in C and D indicate luminal tumors that are likely misclassified.

3.2.4. Developing a miRNA-based PAM classifier

Due to the number of misclassifications using an unbiased miRNA approach, we concluded that a supervised approach based on the mRNA subtype calls would generate more accurate information. Since there is only one publicly available miRNA-seq data set with matched mRNA expression data, we needed to generate a validation cohort. We generated a new miRNA-seq data set by performing small RNA sequencing on 62 FF samples from MDA (11). Using TCGA's cohort as a training set, we used 412 miRNAs that passed filtering criteria in TCGA's cohort and were present in the FF cohort to develop a miRNA-based PAM classifier. PAM identified 63 miRNAs that optimally assigned samples to the basal and luminal subtypes identified by mRNA expression (Figure 9A/B).
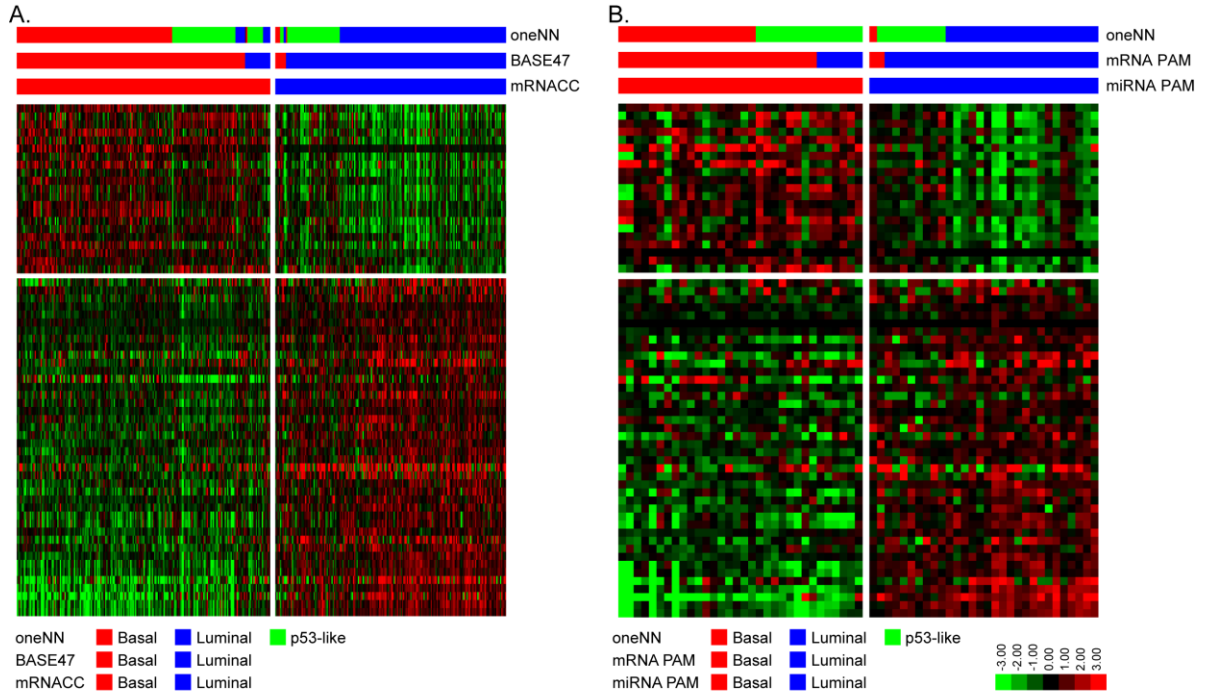
Figure 9: miRNA PAM identifies basal and luminal tumors. A. Visualization of the 63 miRNAs identified by PAM in TCGA's cohort (n=405). B. Visualization of the 63 miRNAs identified by PAM in the FF validation cohort (n=62). Color bars: oneNN – basal (red), luminal (blue), p53-like (green); BASE47 – basal (red), luminal (blue); mRNACC – basal (red), luminal (blue); mRNA/miRNA PAM – basal (red), luminal (blue).

We confirmed that these 63 miRNAs could accurately distinguish basal and luminal tumors using hierarchical clustering (Figure 10). This resulted in 85% and 83% concordance with the subtype assignments made by mRNA consensus clustering and BASE47, respectively.
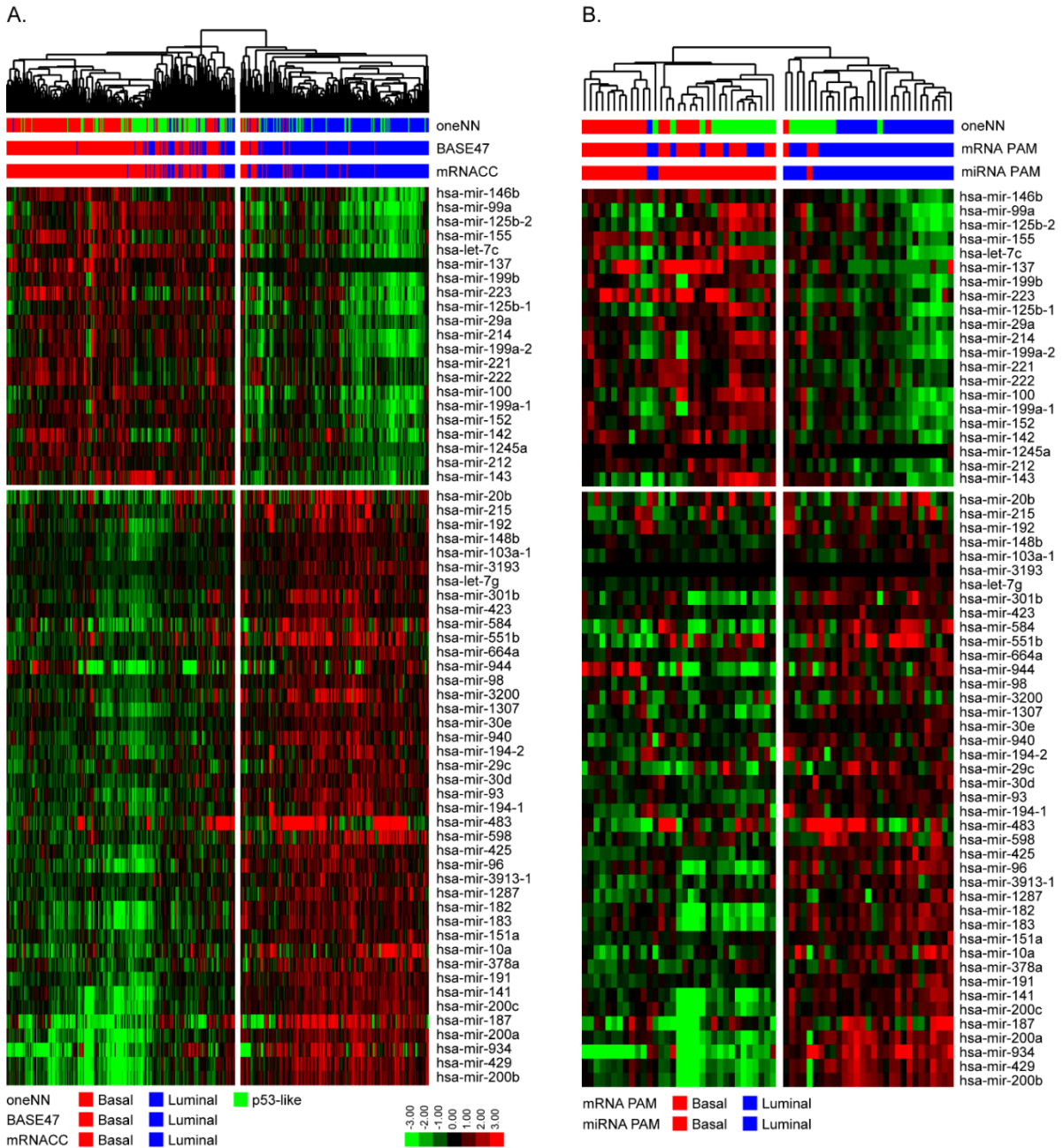
Figure 10: Supervised hierarchical clustering with the 63 PAM miRNAs. A. Hierarchical clustering was performed with the 63 PAM miRNAs in TCGA's cohort, and the relative expression of the 63 miRNAs is depicted in the heatmap. B. Hierarchical clustering was performed with the 63 PAM miRNAs in the FF validation cohort, and the relative expression of the 63 miRNAs is depicted in the heatmap. Color bars: oneNN – basal (red), luminal (blue), p53-like (green); BASE47 – basal (red), luminal (blue); mRNACC – basal (red), luminal (blue); mRNA/miRNA PAM – basal (red), luminal (blue).

3.2.5. miRNA expression patterns relate to basal/luminal biology

A survey of the results suggested that the known biological targets of the miRNAs were relevant to basal and luminal biology. The basal tumors expressed high levels of miR-155, miR-142, miR-221, miR-222, and miR-223, which are miRNAs commonly associated with aggressiveness in other solid tumors and poor prognosis (84-88). The luminal tumors expressed high levels of all the miR-200 family members (miR-200a/b/c, miR-141, miR-429), consistent with previous findings in MIBC (4, 11). Members of the miR-200 family are known inhibitors of EMT that directly target ZEB1 and ZEB2, core EMT transcription factors that directly inhibit transcription of the epithelial adhesion molecule E-cadherin (34). Basal tumors are known to express high levels of ZEB1/2, while the luminal tumors express high levels of E-cadherin and low levels of ZEB1/2 (4, 11). These findings are consistent with the idea that members of the miR-200 family play important roles in controlling the biological properties of the basal and luminal subtypes of MIBC.

3.2.6. miRNA-mRNA relationships and relevance to bladder cancer biology

We used the Ingenuity Pathway Analysis miRNA target filter and miRTarBase to identify additional miRNA-mRNA relationships that were either experimentally observed or were highly predicted in the TargetScan database. The results revealed that many of the basal subtype-associated miRNAs targeted mRNAs involved in adipogenesis, differentiation and EMT suppression (Figure 11A). Specifically, several basal miRNAs (miR-125b, miR-142, miR-143, miR-152, miR-155 and miR-221/222/223) have been predicted to target genes involved in adipogenesis and

RXR activation pathways (89-91). Most notably, miR-125b, miR-223, miR-99a and miR-212 target FGFR2 and FGFR3, which are involved in luminal MIBC biology (4, 91-93). Several basal miRNAs (miR-125b, miR-142, miR-152, miR-146b, miR-222, miR-212) have also been predicted to target luminal factors that have been previously identified in breast cancer, including ERBB2, ERBB3, ERBB4 and FOXA1 (94-96). MiR-125b has also been predicted to inhibit homeobox (HOX) genes, which control urothelial terminal differentiation and are highly expressed by luminal MIBCs (97).

The miRNAs that were enriched in luminal MIBCs were predicted to target pathways associated with invasion and metastasis, including EMT, fibrosis and the actin cytoskeleton (Figure 11B). For example, experimentally observed targets of miR-29c include 6 different collagens that are associated with fibrosis and possibly support cancer-associated fibroblast infiltration (98, 99). MiR-10a, miR-20b and miR-301 were also predicted to target STAT3, while miR-1287 and miR-191 were predicted to target EGFR and IL6, respectively. EGFR and IL6 are upstream regulators of STAT3, all of which have been implicated in basal breast and bladder cancers (100, 101).
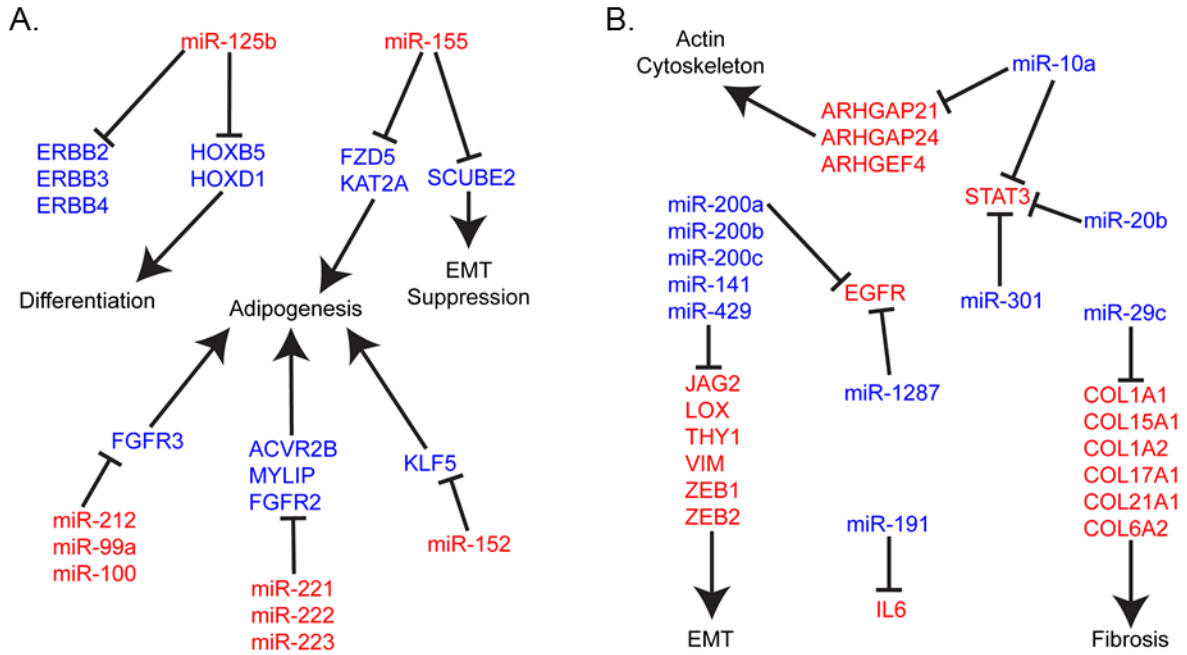
Figure 11: Schematic representation of miRNA and mRNA target relationships.
A. MiRNAs overexpressed in basal tumors (indicated in red text) target luminal
genes (indicated in blue text) involved in differentiation, adipogenesis, and EMT
suppression pathways. B. MiRNAs overexpressed in luminal tumors (indicated in
blue text) target basal genes (indicated in red text) involved in EMT, fibrosis, actin
cytoskeleton, and basal gene expression pathways.

3.2.7. Subsets of basal and luminal MIBC

It was evident from the beginning that a subpopulation of the basal tumors expressed lower levels of both basal and luminal miRNAs, and this subpopulation corresponded to p53-like tumors identified by the oneNN classifier. To explore this heterogeneity further, we utilized the class assignments from the TCGA 3-cluster (k = 3) mRNA CC solution to isolate the significant miRNAs and mRNAs that were differentially expressed by the "infiltrated" CC (Figure 12). We identified 15 miRNAs by differential expression analysis that were overexpressed by the "infiltrated" cluster. We analyzed the basal and luminal subtypes separately, and performed hierarchical clustering with the 15-miRNA signature. In the basal subtype, we isolated 82% of the "infiltrated" tumors identified by mRNA CC (k=3), and in the luminal subtype we isolated 83% of the "infiltrated" tumors (Figure 12A). The "infiltrated" luminal tumors identified by the 15-miRNA signatures likely correspond to TCGA's cluster II, while the "infiltrated" basal tumors show similar expression patterns to TCGA's cluster IV (4).

We applied the 15-miRNA signatures identified in TCGA's cohort to the 62 FF cohort. The basal and luminal tumors as identified by mRNA PAM were analyzed separately, and hierarchical clustering was performed with the 15-miRNA expression signature (Figure 12B). In the basal subtype, all of the "infiltrated" tumors identified by mRNA CC (k=3) and the oneNN p53-like tumors clustered together. In the luminal subtype, 89% of the "infiltrated" tumors identified by CC were isolated by hierarchical clustering. The 15-miRNA signature includes miR-133b, miR-133a,

miR-143, miR-145, miR-99a and miR-100, which have been previously associated

with fibrosis and chemo-resistance (102-104).
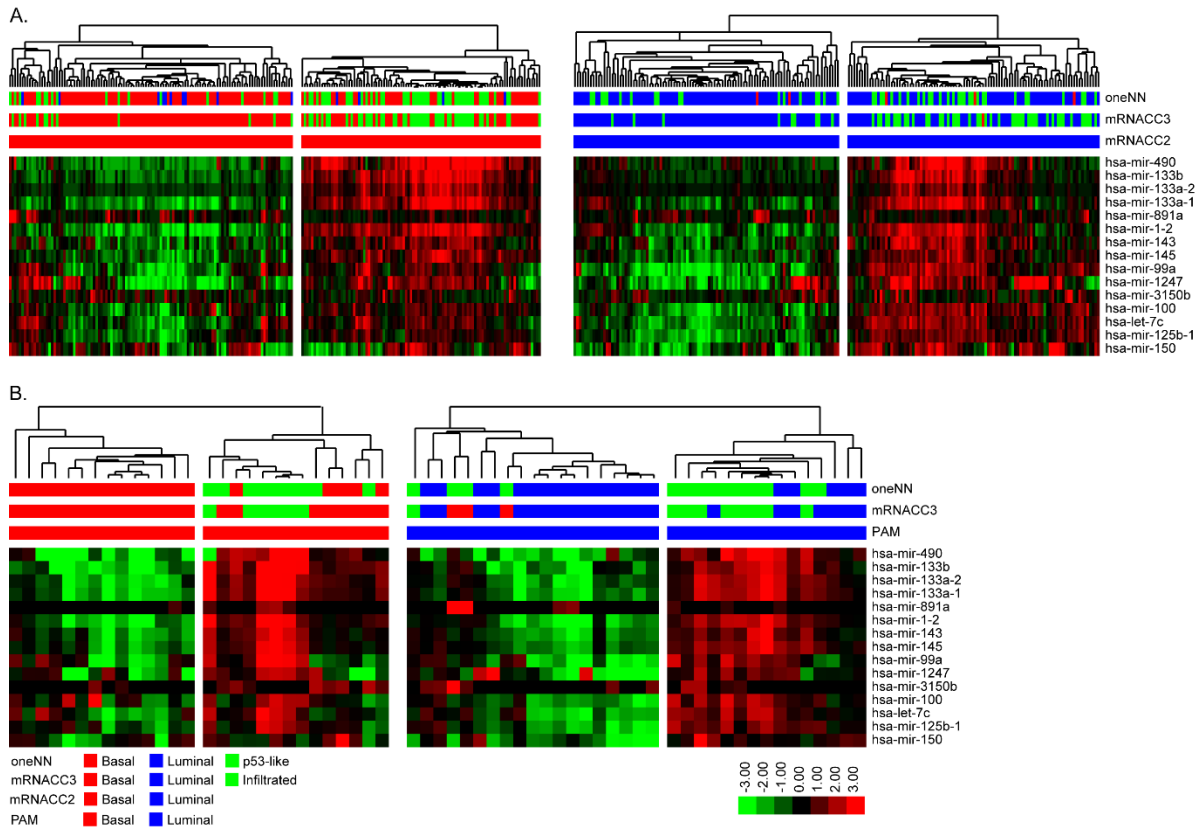
Figure 12: 15-miRNA signature identifies "infiltrated" subsets of the basal and luminal subtypes. A. Hierarchical clustering was performed with the 15-miRNA signature in TCGA's cohort. The heatmaps depict relative expression of the 15-miRNA signature. (Left) Basal tumors identified by mRNA CC were isolated and subjected to supervised hierarchical clustering. (Right) Luminal tumors identified by mRNA CC were isolated and subjected to supervised hierarchical clustering. B. Hierarchical clustering was performed with the 15-miRNA signature in the FF validation cohort. The heatmaps depict relative expression of the 15-miRNA signature. (Left) Basal tumors identified by mRNA PAM were isolated and subjected to supervised hierarchical clustering. (Right) Luminal tumors identified by mRNA PAM were isolated and subjected to supervised hierarchical clustering. Color bars: oneNN – basal (red), luminal (blue), p53-like (green); mRNACC3 – basal (red), luminal (blue), "infiltrated" (green); mRNACC2 – basal (red), luminal (blue); PAM – basal (red), luminal (blue).

59

3.2.8. Survival outcomes

Lastly, we assessed the survival outcomes based on the subtype assignments made using the 63 PAM miRNAs in TCGA's cohort (n=405) to the mRNA based subtype assignments. We compared the two-cluster solutions (BASE47, mRNA CC, miRNA PAM), and we observed that in all cases patients with basal tumors had the poorest clinical outcomes (Figure 13). Having clinically available tests to prospectively identify patients with basal tumors is crucial, as our previous work showed that basal MIBCs responded well to platinum-based chemotherapy (11, 105).
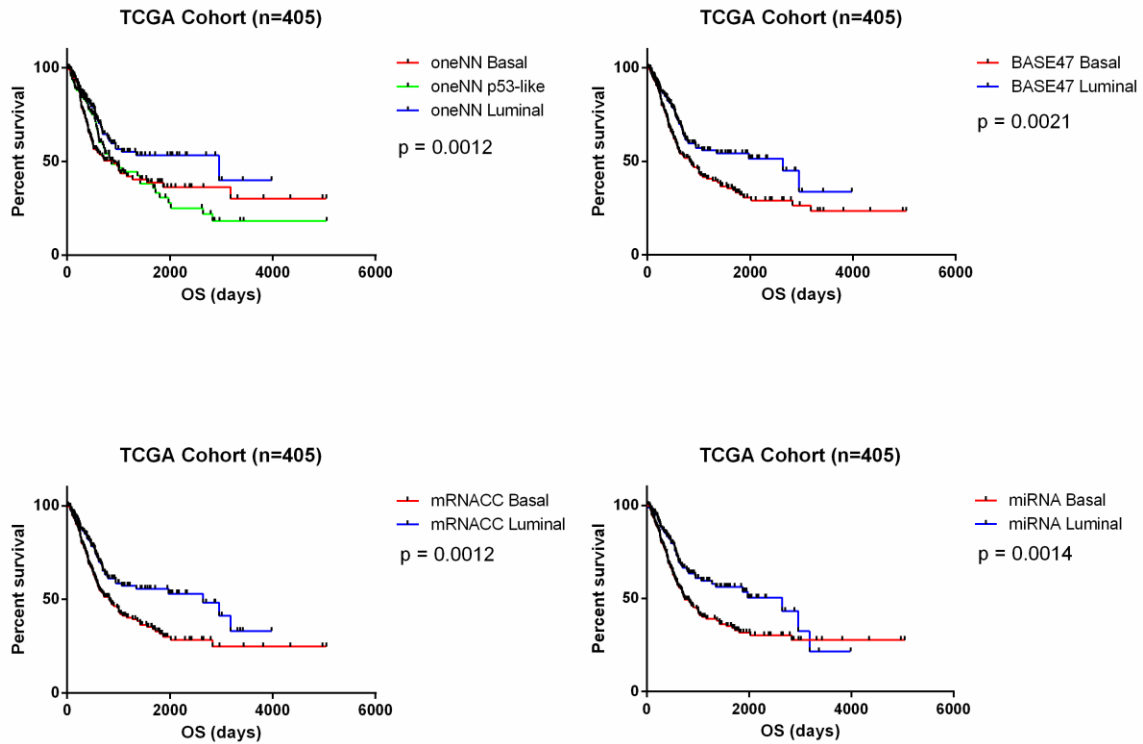
Figure 13: Survival analyses confirm that basal tumors have poor survival outcomes. Survival analyses performed with TCGA's cohort (n=405) confirm that basal tumors identified by four different methods have poor survival outcomes. A. Kaplan-Meier plot of overall survival based on MDA oneNN subtype assignments (p = 0.0012). B. Kaplan-Meier plot of overall survival based on BASE47 subtype assignments (p = 0.0021). C. Kaplan-Meier plot of overall survival based on mRNA CC subtype assignments (p = 0.0012). D. Kaplan-Meier plot of overall survival based on 63-miRNA signature assignments (p = 0.0014).

3.3. Discussion

The first goal of this study was to confirm previous conclusions that had been made regarding the intrinsic subtypes of MIBC (4, 7, 11). Using an unsupervised approach and a high-quality RNA-seq data set, we were able to reproduce our previous conclusion that a 3-subtype solution was mathematically optimal (11). However, one of the subtypes that corresponded to our original p53-like subtype, was unstable (11). Since there were stability issues, we concluded that a two-subtype solution corresponded better with the known biology. We used unsupervised CC, and a k=2 solution, to develop a new mRNA PAM classifier. The subtype identifications made by our new PAM classifier, were directly compared to the assignments made by BASE47 and revealed over 90% concordance. We concluded that the basal and luminal subtypes of MIBC are highly robust, and most likely identify the intrinsic subtypes of bladder cancer.

The second part of this study was to develop a miRNA-based classifier that could be used to assign tumors to the molecular subtypes of MIBC. Since miRNAs are considerably more stable in body fluids, including urine and blood, the development of a miRNA-based classifier present the possibility to perform tumor subtype calls with non-invasive techniques or "liquid biopsies". Also, miRNA expression profiling presents a new opportunity in the assessment of FFPE tissue sections, since the RNA isolated from these sections is often too degraded for sequencing techniques. We used a supervised approach based on mRNA subtype classifications to identify 63 miRNAs that assigned tumors to the basal and luminal subtypes. Analysis of the 63 miRNAs suggested that they may serve as robust

biomarkers, since their biological functions supported previous conclusions made about the biology underlying the basal and luminal subtypes. Most of the overexpressed basal miRNAs targeted mRNAs implicated in the control of luminal biology (FOXA1, ERBB2), adipogenesis (FGFR2, FGFR3), and urothelial differentiation (HOX genes), which are down-regulated in the basal subtype. The miRNAs overexpressed in the luminal subtype targeted transcription factors that control EMT (ZEB1/2), biomarkers associated with fibrosis and the actin cytoskeleton (collagens), and basal cancer biology (IL6, EGFR, STAT3), which are all suppressed in luminal cancers. Using the 63-miRNA signature, were able to accurately assign MIBCs to the basal and luminal subtypes, and confirmed that the basal tumors were associated with poor outcomes. We also identified a 15-miRNA signature that identified sub-populations within the basal and luminal subtype that appeared to be infiltrated with fibroblasts. This signature could prove to be useful in identifying tumors that are resistant to traditional chemotherapy (11, 102, 106, 107), but are sensitive to immune checkpoint blockade (6).

Overall, we identified high concordance between the calls made using BASE47 and the new mRNA PAM classifier described here, demonstrating that subtype identifications made using relative mRNA expression are highly robust. It is becoming common practice to sequence the tumors of every single patient. Given the number of currently available subtype classifiers, it should become common clinical practice to subtype the tumors. However, this can be difficult when using FFPE tissues, because of the variability in the quality of the RNA isolated. Often times, samples are not usable because they do not pass standard quality control

cut-offs. The identification of a miRNA-based classifier presents new opportunities

for subtype classification, since miRNAs are often stable in FFPE tissues and

degraded RNA samples. While the miRNA-based classifier does not appear to be

as robust as an mRNA-based classifier, it could be used as an alternative strategy

when samples do not pass quality checks. A miRNA-based classifier could also be

paired with an immunohistochemical classifier to confirm the subtype classifications

(108). Overall, the clinical identification of tumor subtype based on mRNA, miRNA,

or immunohistochemistry should become part of routine diagnostic practices for

MIBC in the near future.

Chapter 4: Results

LncRNA expression subtypes in MIBC

Chapter 4: LncRNA expression subtypes in MIBC

4.1. Introduction and Rationale

Our understanding of cancer biology has transformed recently with the latest, large-scale cancer genomics projects that have been done. Recent studies in MIBC have catalogued common DNA alterations that are present in MIBCs, and several groups also employed whole genome mRNA expression profiling to identify intrinsic basal and luminal subtypes of MIBCs that are similar to the corresponding subtypes of breast cancer (4, 7, 8, 11). While several studies have focused on the impact of DNA mutations, copy number variations and translocations, non-protein coding RNAs have been overlooked on the large scale.

Non-protein coding RNAs, including miRNAs and lncRNAs, have been shown to have a greater impact on evolutionary diversity than protein-coding mRNAs (109, 110). The effects of miRNAs on cancer biology have been studied, and their primary function is widely understood to inhibit expression of proteins encoded by their mRNA targets. However, lncRNAs are a little more of a mystery since they are not as well understood, but have been observed to contribute to tumor initiation and progression (111-113). While lncRNAs have been implicated in various biological functions, the mechanisms by which they exert their activity is largely unknown. Until recently, whole genome approaches to measure lncRNA expression patterns were not available, which has limited the scope of investigations to date. In this chapter, we developed bioinformatic methods to extract whole genome lncRNA expression patterns from the TCGA's RNA-seq data to examine whether we could

also identify lncRNA-based intrinsic subtypes of MIBC. We seek to not only deepen

our current understanding of MIBC biology, but to also identify potential biomarkers

to clinically characterize the disease.

4.2. Results

4.2.1. Identification of lncRNA-based MIBC subtypes

We examined patterns of whole-genome lncRNA expression in TCGA's MIBC

RNA-seq dataset (n=407) by re-aligning and re-annotating the data. A total of

31,191 lncRNAs were isolated and analyzed. After filtering based on level of

expression, 1,827 lncRNAs were found to be expressed by MIBCs. We used the

1,827 lncRNAs to perform unsupervised consensus clustering (CC) to determine the

number of intrinsic lncRNA-associated molecular subtypes that were present. The

results revealed the presence of 3 distinct clusters (Figure 14A). We compared the

lncRNA subtypes to the mRNA based basal, luminal, and p53-like subtypes as

identified by MDA's oneNN model (11), and found 75% concordance. We noticed

that the third cluster identified by CC contained mixed expression of the basal and

luminal specific mRNAs (Figure 14B), similar to our previous findings (114),

suggesting a 2-cluster solution may be more robust.

A.
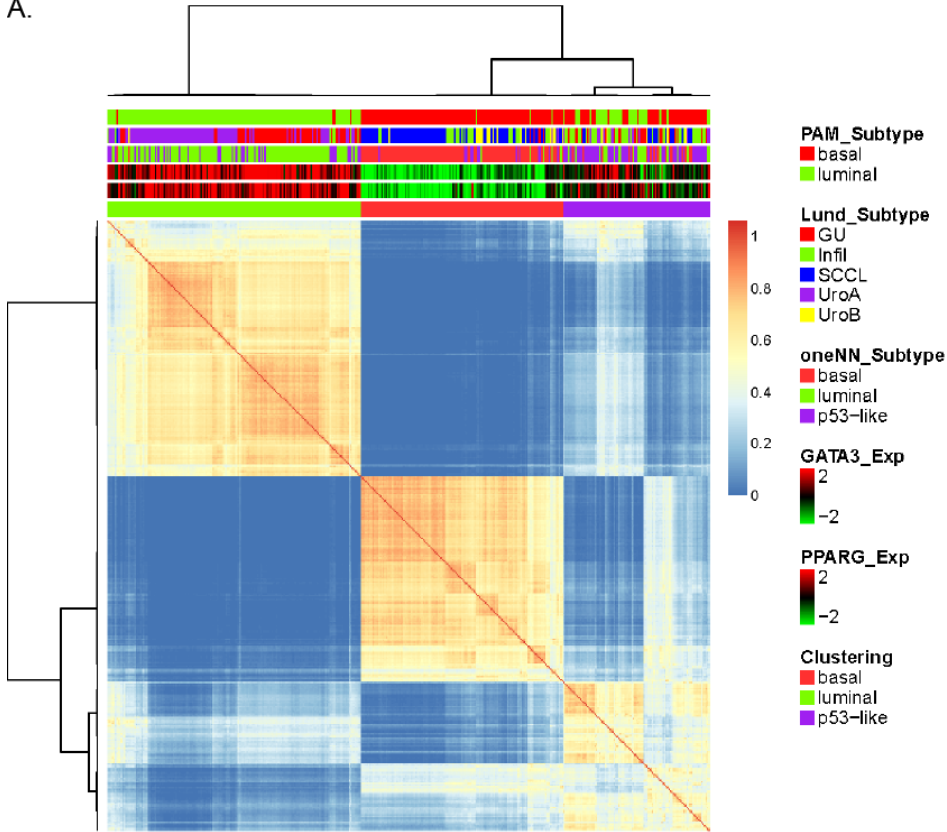
B.
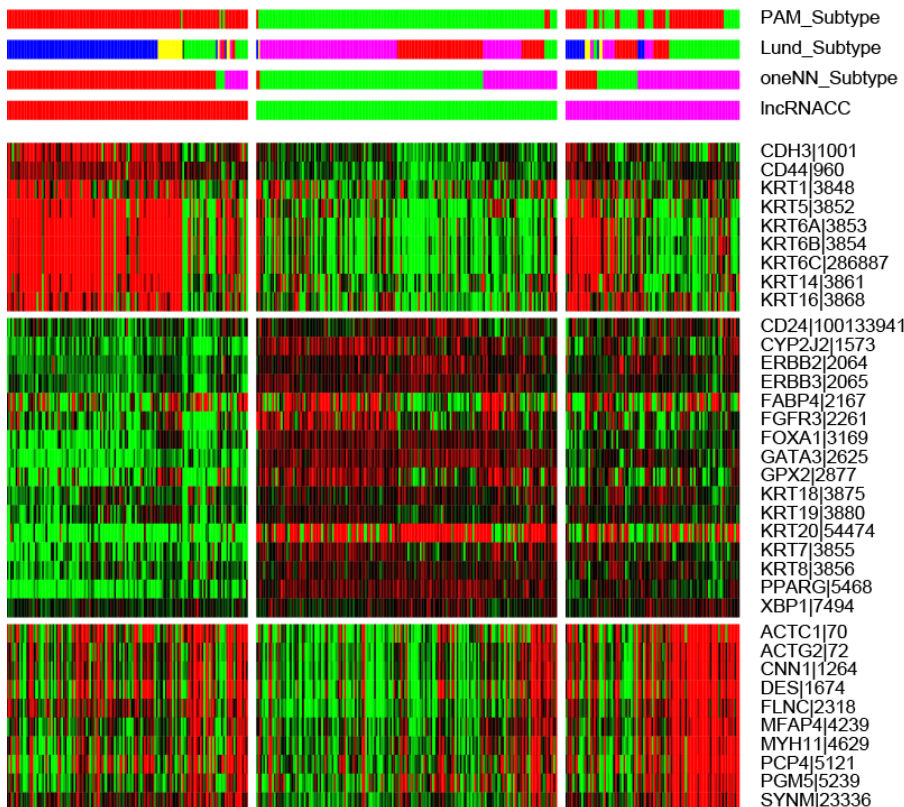
Figure 14: LncRNA consensus clustering identifies a 3-cluster solution.
LncRNA consensus clustering was performed with TCGA's cohort (n=407) and a 3-cluster solution was identified as mathematically optimal. A. Consensus clustering results in 3 distinct clusters. B. Visualization of MDA basal markers (top), MDA luminal markers (middle), and MDA p53-like tumors (bottom). Color bars: PAM subtype – basal (red), luminal (green); Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); oneNN subtype – basal (red), luminal (green), p53-like (purple); lncRNA consensus clustering – basal (red), luminal (green), p53-like (purple). Relative expression of luminal genes GATA3 and PPARG are also visualized.

We chose to focus on a 2-cluster solution, focusing on the intrinsic basal and luminal subtypes, since these appeared to be more stable (Figure 15). When the lncRNA-based basal and luminal subtypes were compared to the mRNA PAM-identified basal and luminal subtypes (114), there was 98% concordance, suggesting that lncRNA expression alone could be used to identify the intrinsic basal and luminal subtypes of MIBC. We also compared the lncRNA CC subtype identifications to the oneNN basal and luminal classifications, in the absence of the p53-like tumors (11), and we also see 98% concordance between the calls, suggesting the basal and luminal intrinsic subtypes are reproducible.

A.
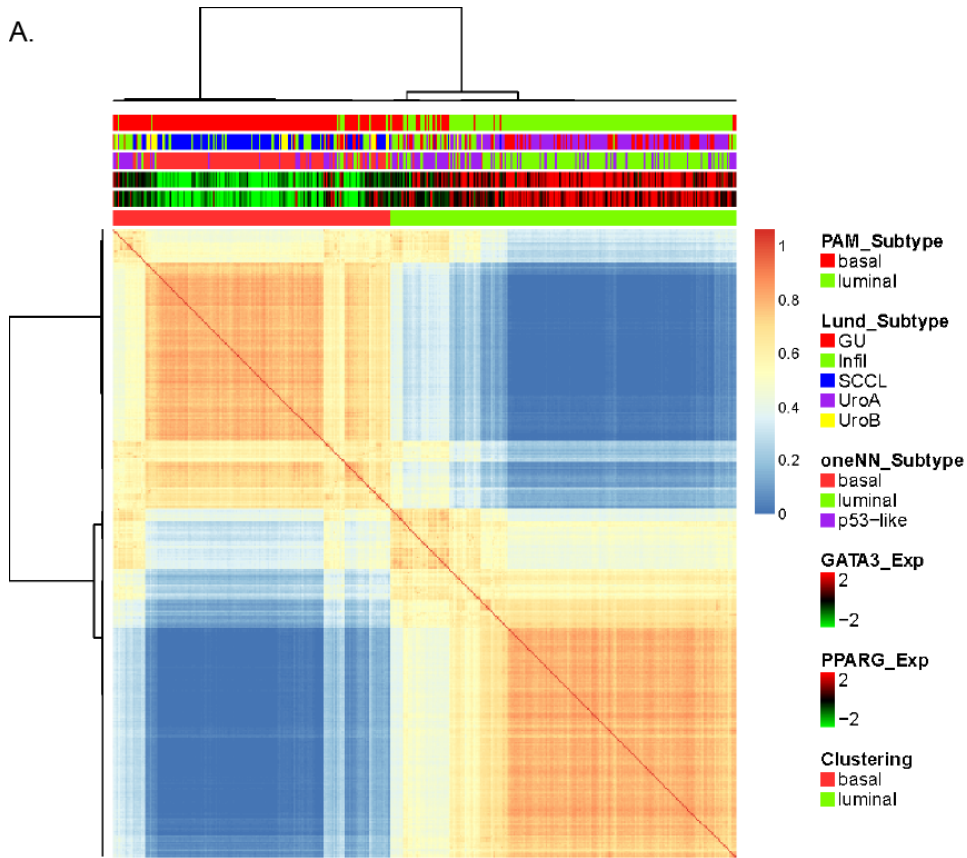
B.
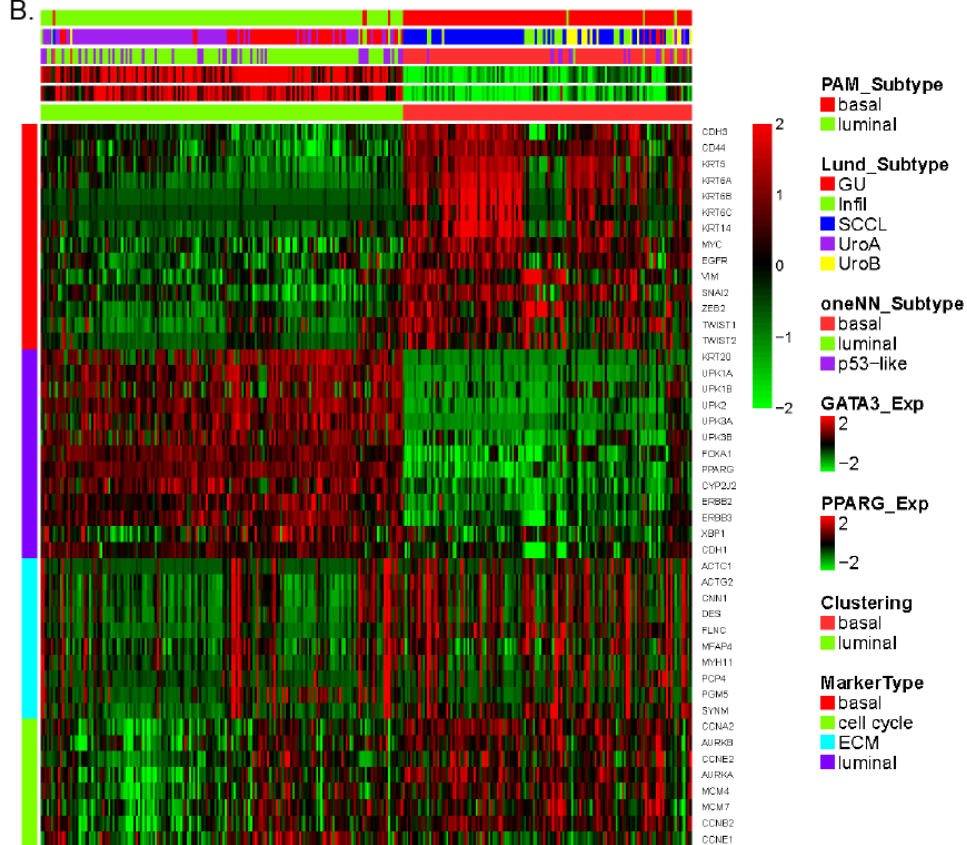
Figure 15: 2-cluster lncRNA consensus clustering solution. LncRNA consensus clustering was performed with TCGA's cohort, in the absence of "p53-like" tumors (n = 308) and a 2-cluster solution was identified. A. LncRNA consensus clustering results in 2 distinct clusters. B. Visualization of basal markers (top), cell cycle markers (middle - top), extracellular matrix markers (middle - bottom), and luminal markers (bottom). Color bars: PAM subtype – basal (red), luminal (green); Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); oneNN subtype – basal (red), luminal (green), p53-like (purple); lncRNA consensus clustering – basal (red), luminal (green). Relative expression of luminal genes GATA3 and PPARG are also visualized.

4.2.2. mRNA expression in the lncRNA-defined subtypes

We used the mRNA expression patterns in the lncRNA-defined basal and luminal MIBCs to perform gene set enrichment analysis (GSEA). The results indicated that the basal and luminal MIBC clusters identified by lncRNAs were very similar to basal and luminal breast cancers, as indicated by the significant enrichment of the CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_UP gene signature in the luminal subtype and CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_DOWN gene signature in the basal subtype (115). Also, similar to previous reports in bladder and breast cancers, the basal subtype identified by lncRNA expression had overall poor survival when compared to the luminal subtype (4, 7, 11, 114) (Figure 16).
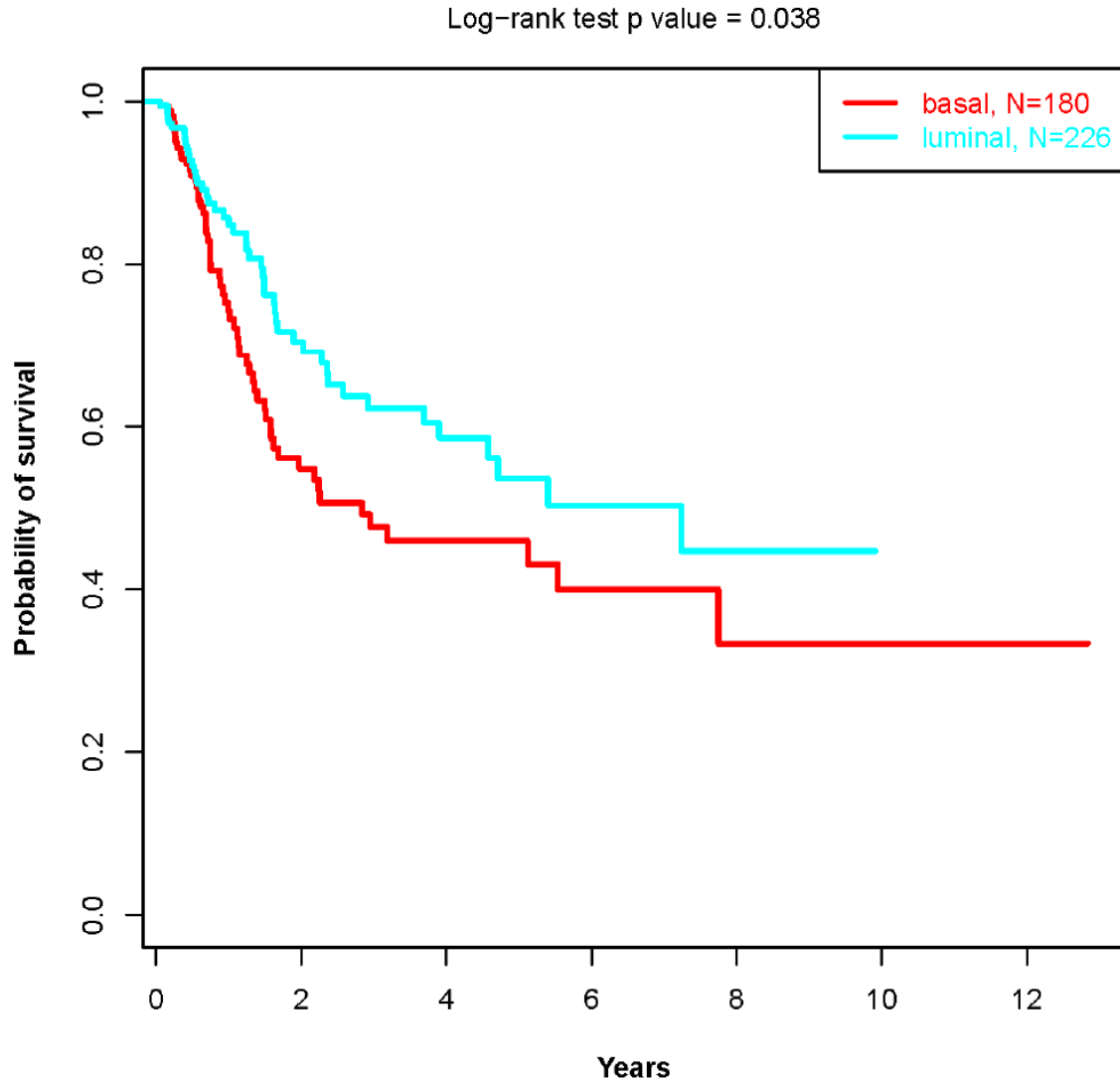
Figure 16: Survival outcomes based on the lncRNA defined basal and luminal clusters. Kaplan-Meier plot of the overall survival outcomes of the lncRNA defined basal and luminal clusters in TCGA's cohort (n = 308). The basal subtype exhibited overall poor survival outcomes (p = 0.038).

4.2.3. Differential lncRNA expression

Differential expression analysis of the lncRNAs between the basal and luminal subtypes identified 430 lncRNAs to be significantly differentially expressed (Figure 17). The basal subtype overexpressed 204 lncRNAs, while the luminal subtype overexpressed 226 lncRNAs. A closer analysis of the differentially expressed lncRNAs identified some lncRNAs with previously established functional activity in cancer. The basal subtype overexpressed several lncRNAs that have been previously studied, including MIAT, MEG3, and MIR155HG, while the luminal subtype overexpressed UCA1. Interestingly, the basal subtype expressed MIR155HG, which has been associated with an aggressive phenotype and activation of epithelial-mesenchymal transition (EMT) (116). There was also high expression of MEG3, which has been previously characterized as a tumor suppressor in bladder cancer since it is typically expressed by normal tissues and lost in bladder cancer (69, 117). Urothelial carcinoma associated 1 (UCA1) has been identified as a biomarker for bladder cancer in urine sediment (65), and in this study, we found it to be significantly upregulated in luminal MIBC when compared to basal MIBCs.
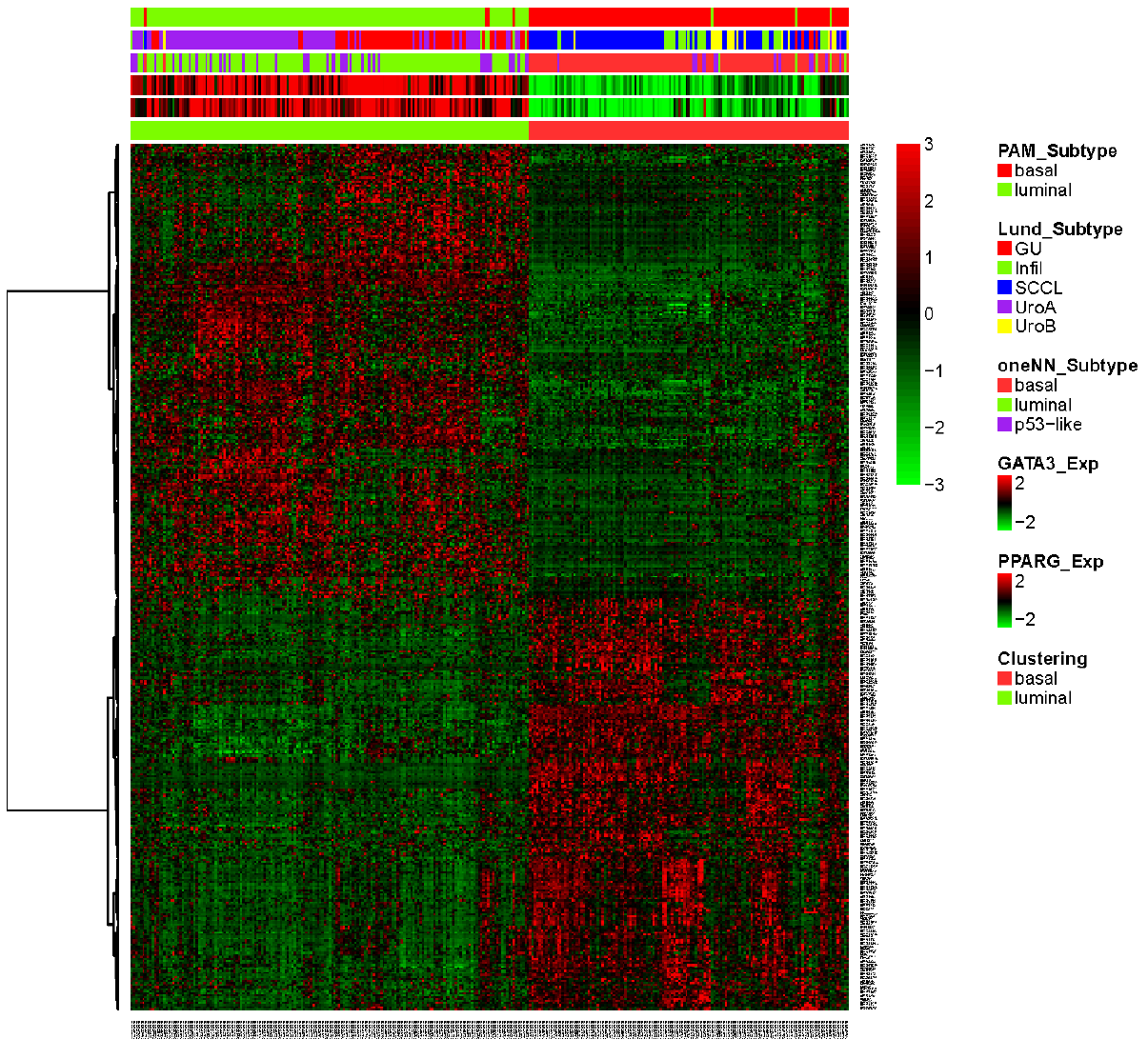
Figure 17: Visualization of the 430 differentially expressed lncRNAs. The heatmap depicts relative expression of the 430 significantly differentially expressed lncRNAs in TCGA's cohort (n = 308). Color bars: PAM subtype – basal (red), luminal (green); Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); oneNN subtype – basal (red), luminal (green), p53-like (purple); lncRNA consensus clustering – basal (red), luminal (green). Relative expression of luminal genes GATA3 and PPARG are also visualized.

4.2.4. LncRNA-based PAM classifier

We next sought to develop a lncRNA-based PAM classifier, which could accurately identify the basal and luminal subtypes of MIBC. Using TCGA's cohort and the lncRNA CC subtype identifications, PAM identified solutions ranging from 53 (Δ = 8.678, overall error rate = 0.013) to 9 lncRNAs (Δ = 12.475, overall error rate = 0.045) that would be optimal for assigning cluster membership (Figure 18). However, due to a lack of publicly available MIBC RNA-sequencing cohorts, the lncRNA-based PAM classifier could not be validated.

lncRNACC ■ Basal ■ Luminal
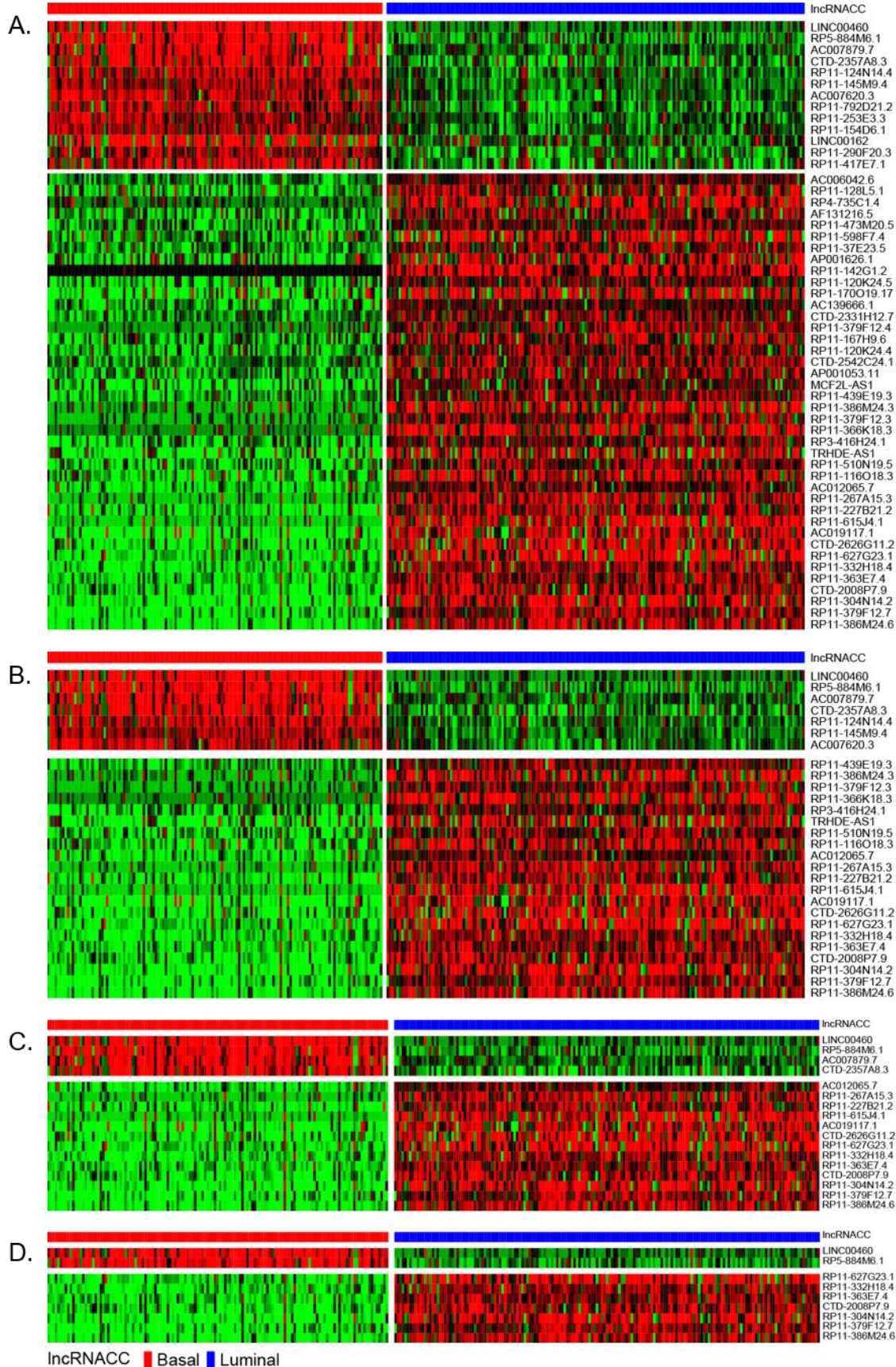
Figure 18:   LncRNA-based PAM classifier training.   A. PAM training solution with 53 lncRNAs (Δ = 8.678, overall error rate = 0.013). B. PAM training solution with 28 lncRNAs (Δ = 9.763, overall error rate = 0.016). C. PAM training solution with 17 lncRNAs (Δ = 10.848, overall error rate = 0.023). D. PAM training solution with 9 lncRNAs (Δ = 12.475, overall error rate = 0.045). Color bar: lncRNACC - basal (red), luminal (blue).

4.2.5. Identification of sub-clusters within the basal and luminal subtypes

We further examined the presence of sub-clusters within the intrinsic basal and luminal lncRNA subtypes. To study this, we performed unsupervised consensus clustering on the 137 basal tumors identified by lncRNA CC. This revealed the presence of 2 clusters (Figure 19A). We used canonical basal markers (KRT5, KRT6A, KRT6B, KRT6C) and markers of EMT (ZEB1/2, TWIST1, FN1) to classify the tumors as either "epithelial" or "mesenchymal" (Figure 20). Comparison of the sub-clusters as identified by mRNA signature or lncRNA CC revealed 90% concordance. There were 129 lncRNAs significantly differentially expressed between the two basal sub-clusters, 62 over-expressed by cluster2 ("epithelial"), and 67 over-expressed in cluster1 ("mesenchymal") (Figure 19B). When comparing the two basal sub-clusters, there was a significant difference in survival outcomes (p=0.027), with the "mesenchymal" tumors showing poor outcomes (Figure 19C).
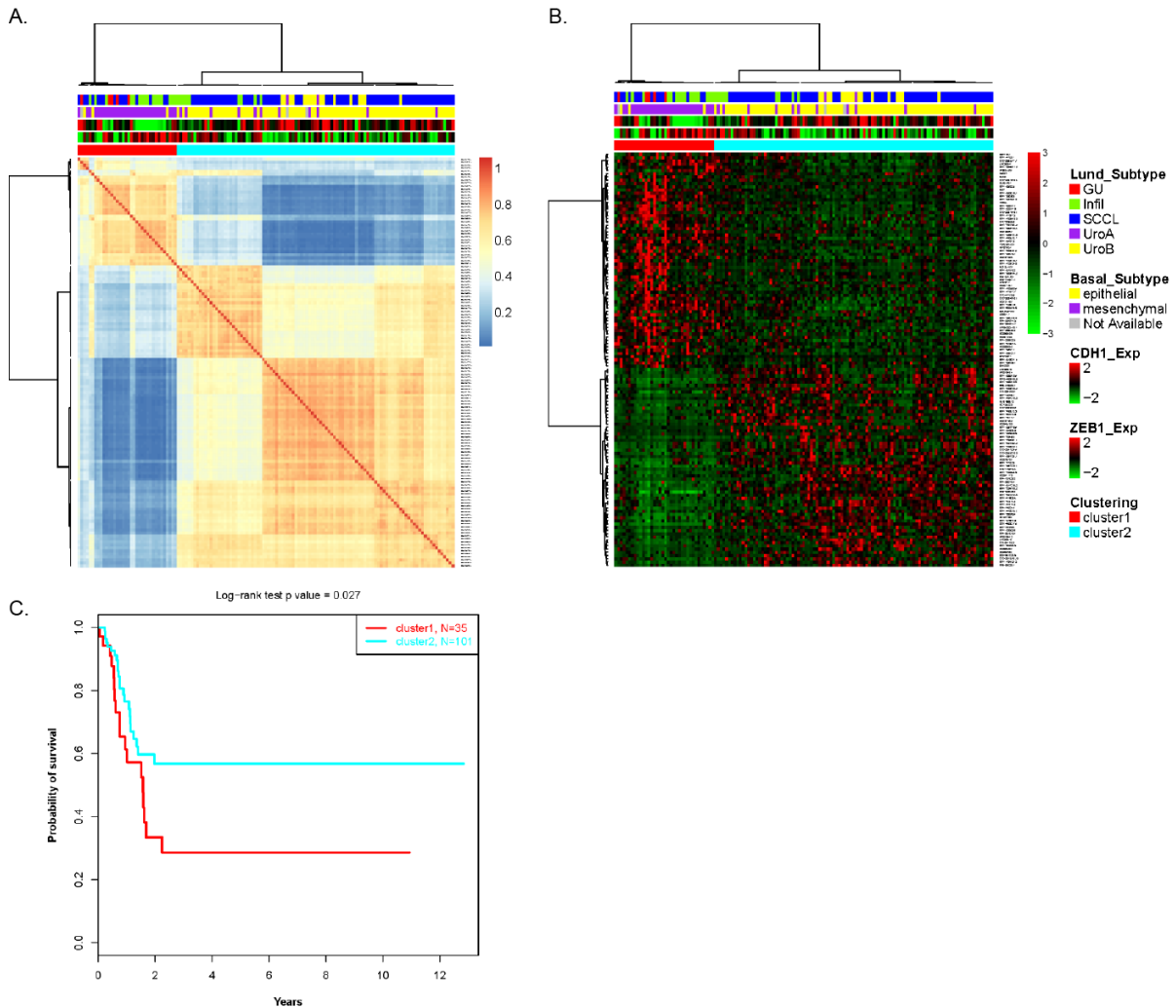
Figure 19: LncRNA consensus clustering identifies two subsets within the basal subtype. The basal tumors identified by lncRNA consensus clustering were subjected to a second round of unsupervised consensus clustering. A. Consensus clustering identifies two subsets within the basal subtype. B. Visualization of the 129 significantly differentially expressed lncRNAs between the two subsets. C. Kaplan-Meier plot shows that cluster1 ("mesenchymal" sub-cluster) exhibits poor overall survival outcomes (p = 0.027). Color bars: Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); basal subtype – epithelial (yellow), mesenchymal (purple), gray (data not available); lncRNA consensus clustering – cluster1 (red), cluster2 (light blue). Relative expression of epithelial marker CDH1, and mesenchymal marker ZEB1 are also visualized.
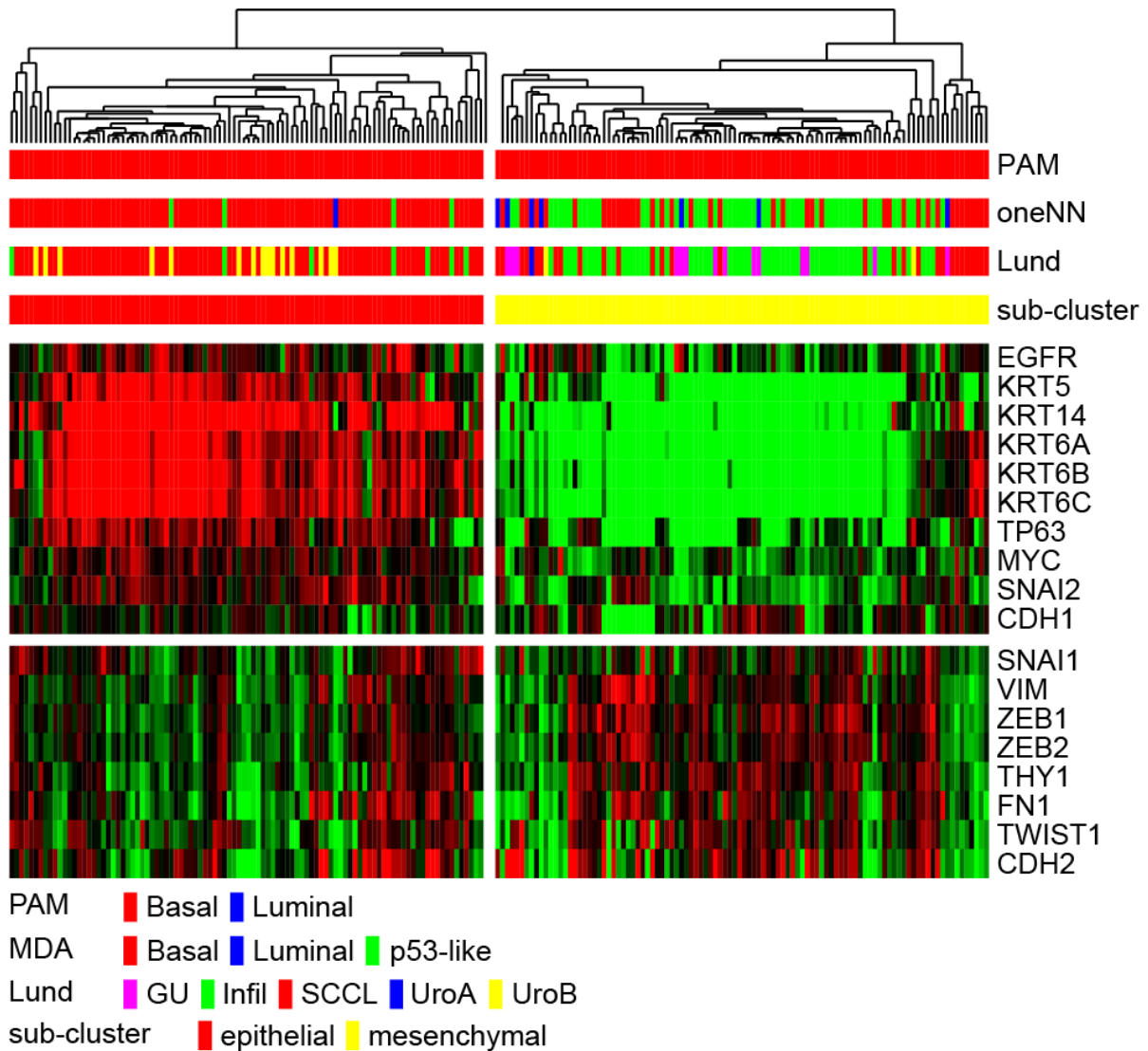
Figure 20: Identification of basal sub-clusters by mRNA expression signature.
Basal tumors in TCGA's cohort of 408 patient samples were isolated using a PAM classifier based on mRNA expression. Supervised hierarchical clustering with epithelial markers (top) and mesenchymal markers (bottom) identified two sub-clusters of basal tumors. Color bars: PAM subtype – basal (red); Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); oneNN subtype – basal (red), luminal (green), p53-like (purple); sub-cluster – epithelial (red), mesenchymal (yellow).

GSEA comparing the mRNA expression patterns between the two sub-clusters confirmed a down-regulation of metastasis markers in the "epithelial" sub-cluster (Figure 21). We found the

CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_UP gene set (Figure 21A) (115), and the JAEGER_METASTASIS_DN gene set (Figure 21B) (118) to be significantly enriched in the "epithelial" basal subset. We followed up this observation by generating an EMT lncRNA expression signature that was derived from cell line expression profiles that are currently available through the Cancer Cell Line Encyclopedia (CCLE). Visualization of this signature showed relatively high expression of EMT lncRNAs in the mesenchymal sub-cluster (Figure 22A), and GSEA analyses showed significant enrichment of the down-regulated lncRNAs in mesenchymal cell lines in the "epithelial" sub-cluster (p = 0, FDR = 0) (Figure 22B). Interestingly, the "mesenchymal" sub-cluster had over-expression of two lncRNAs that are considered to be tumor suppressors, MIAT and MEG3 (60, 69, 117, 119), while the lncRNAs over-expressed in the "epithelial" sub-cluster have not been studied previously. Again, we observed high expression of MEG3 was associated with poor overall survival.
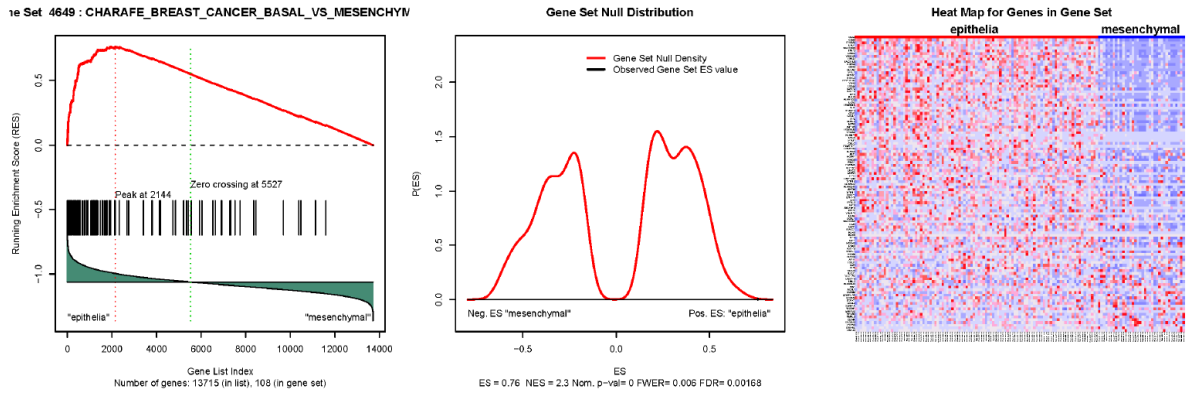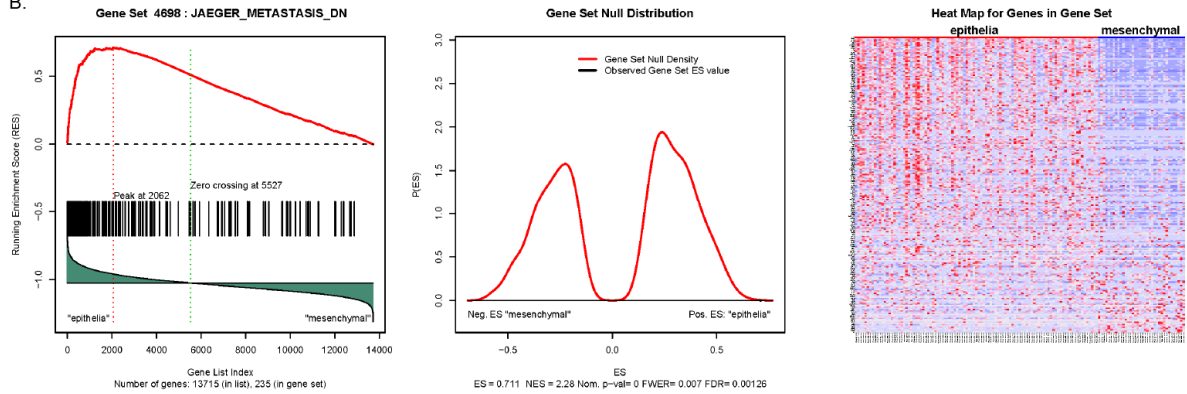
Figure 21: The two subsets of the basal subtype are characterized by differences in expression of mesenchymal gene sets. A. GSEA analyses identified significant enrichment of the CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_UP gene set in the "epithelial" subset (p = 0, FDR = 0.00168). B. GSEA analyses identified significant enrichment of the JAEGER_METASTASIS_DN gene set in the "epithelial" subset (p = 0, FDR = 0.00126).
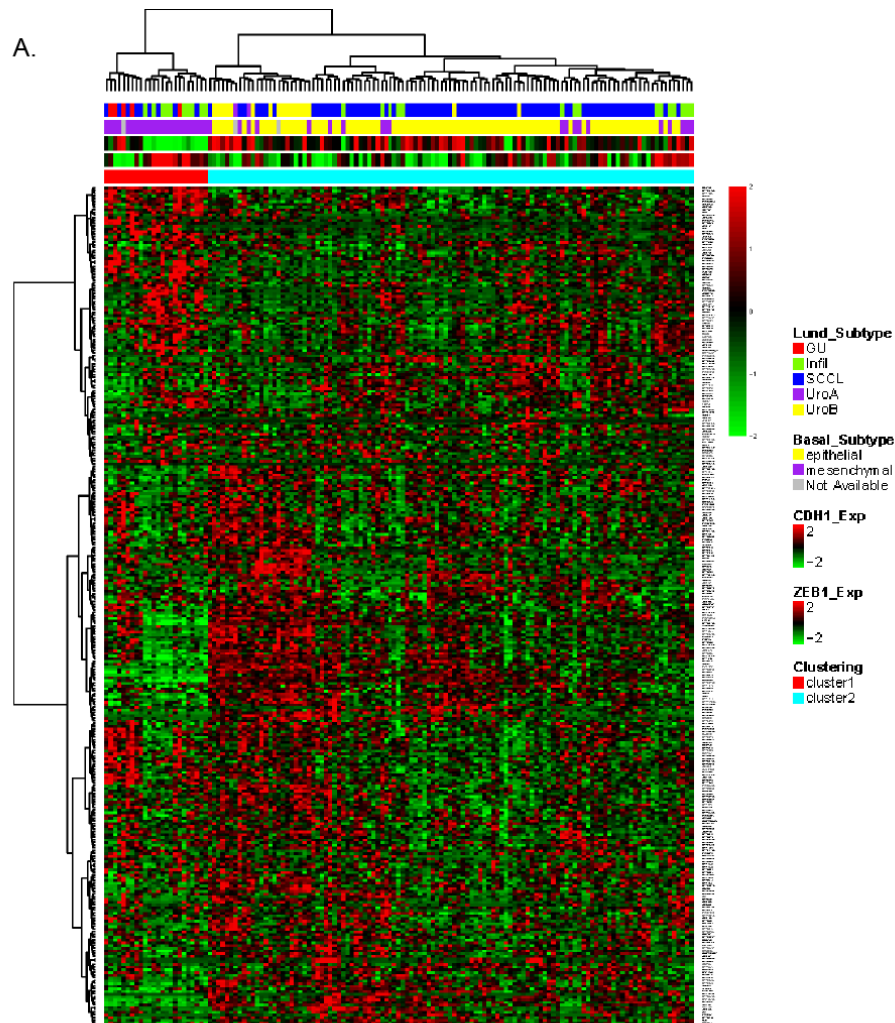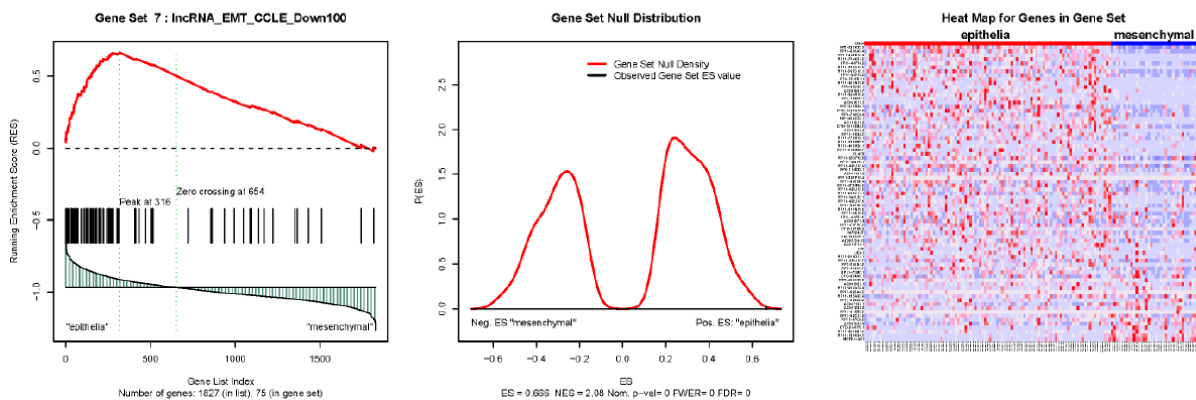
Figure 22: EMT lncRNA expression signature in the basal subtype. A. Visualization of the EMT lncRNA expression signature in the basal subtype identified by lncRNA expression. B. GSEA analyses with a signature of significantly down-regulated lncRNAs in mesenchymal cell lines, showed significant enrichment in the

"epithelial" sub-cluster (p = 0, FDR = 0). Color bars: Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); basal subtype – epithelial (yellow), mesenchymal (purple), gray (data not available); lncRNA consensus clustering – cluster1 (red), cluster2 (light blue). Relative expression of epithelial marker CDH1, and mesenchymal marker ZEB1 are also visualized.

We performed the same unsupervised consensus clustering on the 171 luminal tumors identified by lncRNA CC. This revealed the presence of 2 clusters (Figure 23). We opted to compare these two sub-clusters to the luminal subtypes previously identified by the Lund group, genomically unstable (GU) and urobasal A (UroA) (8). Surprisingly, we found that one of the luminal sub-clusters contained only UroA tumors (cluster2), while the other sub-cluster contained a majority GU tumors along with some UroA tumors (cluster1). We studied the differentially expressed lncRNAs between the two luminal sub-clusters, resulting in 191 significant differentially expressed lncRNAs. There were 58 up-regulated in the "UroA" sub-cluster, and 133 up-regulated in the "GU" sub-cluster (Figure 23B). The "UroA" sub-cluster was enriched for MIR31HG and UCA1, while the "GU" sub-cluster was enriched for DNM3OS, SNHG18, and CDKN2B-AS1. When looking at the two sub-clusters of luminal, there was no difference in survival outcomes (Figure 23C).

GSEA comparing the two sub-clusters of the luminal subtype identified a significant difference in grade based on the presence of the LINDGREN_BLADDER_CANCER_CLUSTER_3_DN gene set in the "UroA" tumors, and the LINDGREN_BLADDER_CANCER_CLUSTER_3_UP gene set in the "GU" tumors (120). This suggested that the "UroA" sub-cluster was enriched with down-regulated genes in high grade tumors (Figure 24A), and the "GU" sub-cluster was enriched with up-regulated genes in high grade tumors (Figure 24B).
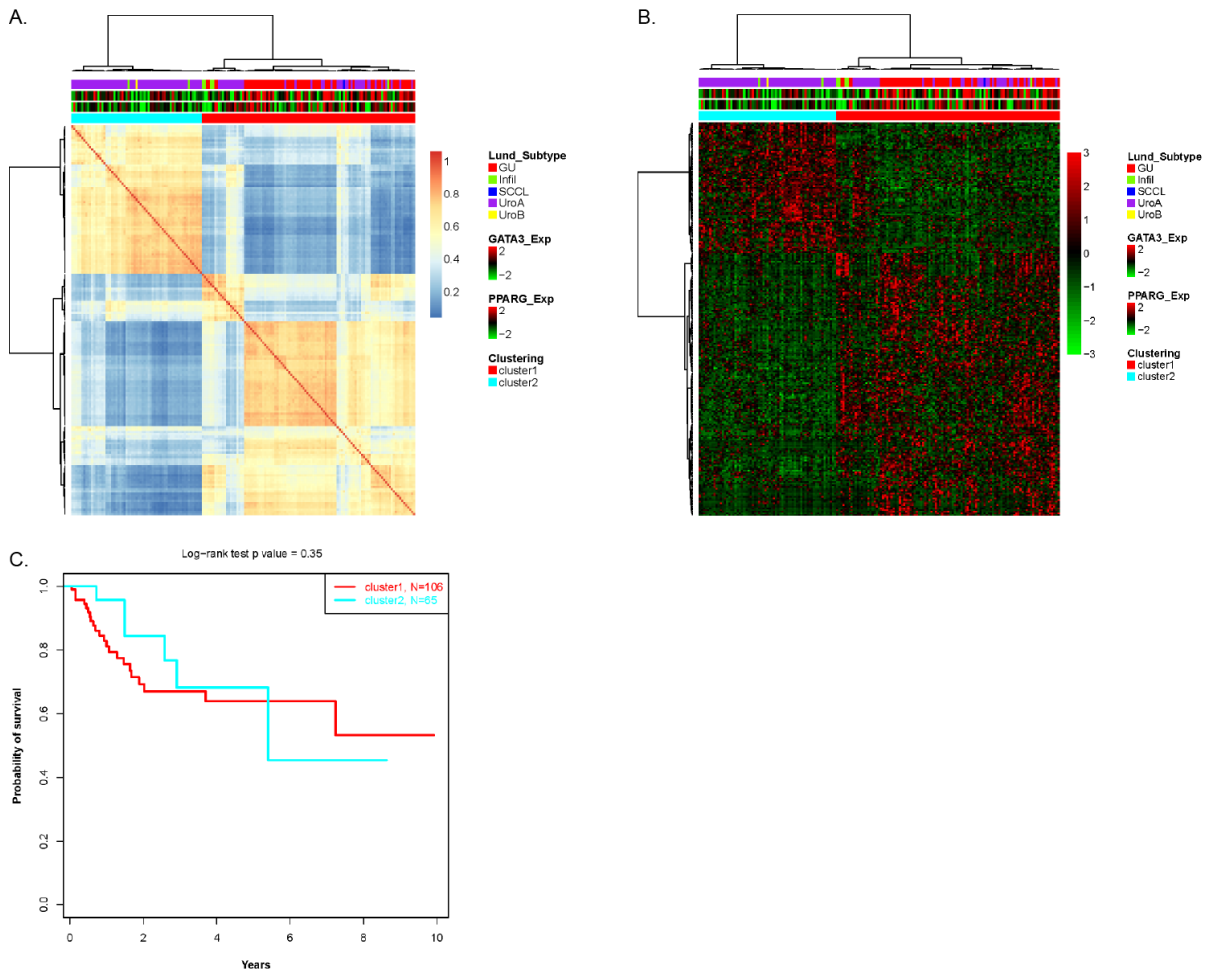
Figure 23: LncRNA consensus clustering identifies two subsets within the luminal subtype. The luminal tumors identified by lncRNA consensus clustering were subjected to a second round of unsupervised consensus clustering. A. Consensus clustering identifies two subsets within the luminal subtype. B. Visualization of the 191 significantly differentially expressed lncRNAs between the two subsets. C. Kaplan-Meier plot shows no significant survival differences between the two subsets (p = 0.35). Color bars: Lund subtype – GU (red), Infil (green), SCCL (blue), UroA (purple), UroB(yellow); basal subtype – epithelial (yellow), mesenchymal (purple), gray (data not available); lncRNA consensus clustering – cluster1 (red), cluster2 (light blue). Relative expression of luminal markers GATA3 and PPARG are also visualized.
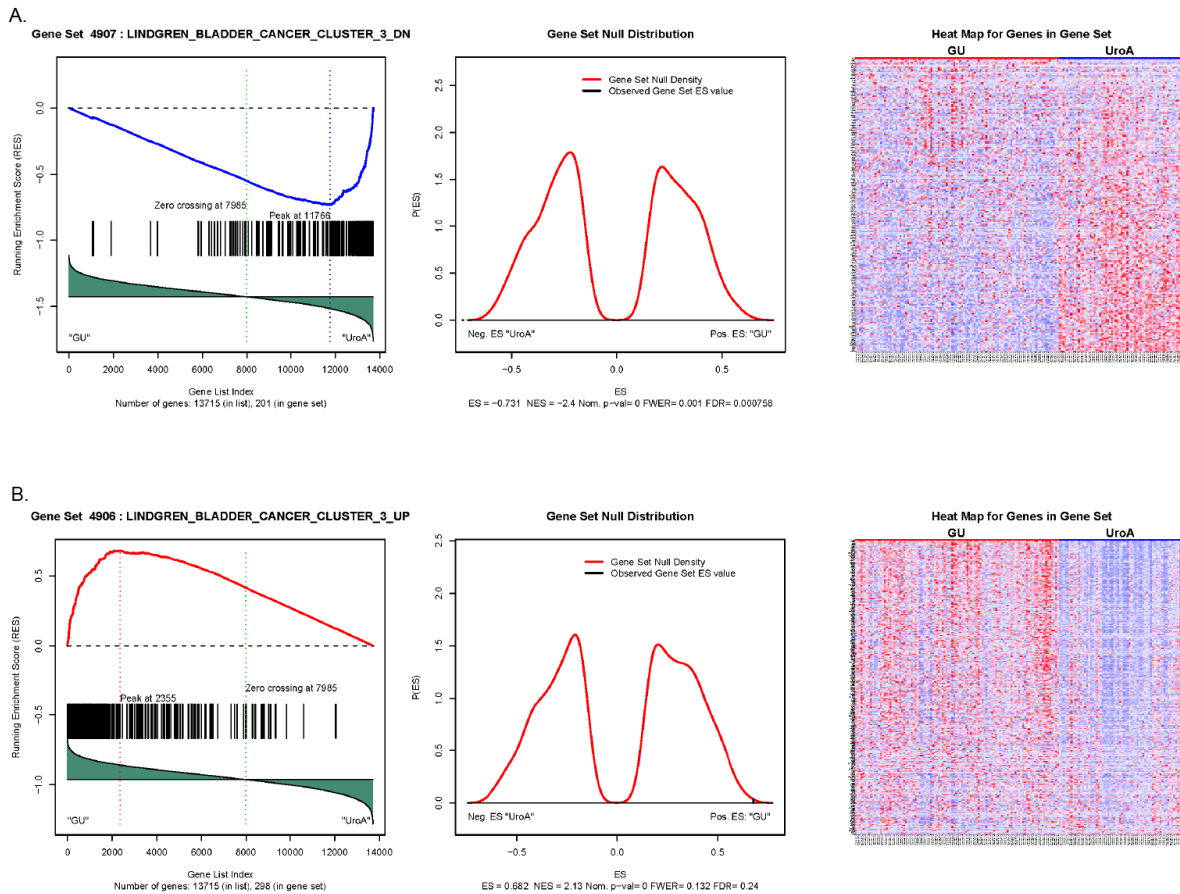
Figure 24: The two subsets of the luminal subtype are characterized by differences in expression of high-grade bladder cancer gene sets. A. The LINDGREN_BLADDER_CANCER_CLUSTER_3_DN gene set is significantly enriched in the "UroA" tumors, suggesting that the "UroA" subset is enriched with tumors of a lower grade (p = 0, FDR = 0). B. The LINDGREN_BLADDER_CANCER_CLUSTER_3_UP gene set is enriched in the "GU" tumors (p = 0, FDR = 0.24), suggesting that the "GU" subset is enriched with high grade tumors.

4.2.6. Summary of differentially expressed lncRNAs

We summarized the findings and identified 40 lncRNAs that were specific to the basal and luminal intrinsic subtypes, as well as to the sub-clusters identified within each intrinsic subtype (Figure 25). The top 10 lncRNAs identified to be overexpressed in the "UroA" sub-cluster were relatively expressed in the other luminal sub-cluster, but were down-regulated in the two basal clusters. The lncRNAs specific to the "GU" sub-cluster exhibited a similar pattern of expression, with slight expression in the other luminal sub-cluster and down-regulation in the two basal clusters. The lncRNAs overexpressed by the basal "epithelial" sub-cluster were significantly down-regulated in the basal "mesenchymal" sub-cluster, and exhibited relatively low expression across the other subsets. While the lncRNAs specific to the basal "mesenchymal" sub-cluster were expressed by the basal "epithelial" sub-cluster to a slightly lower degree, and were relatively down-regulated in the luminal clusters.
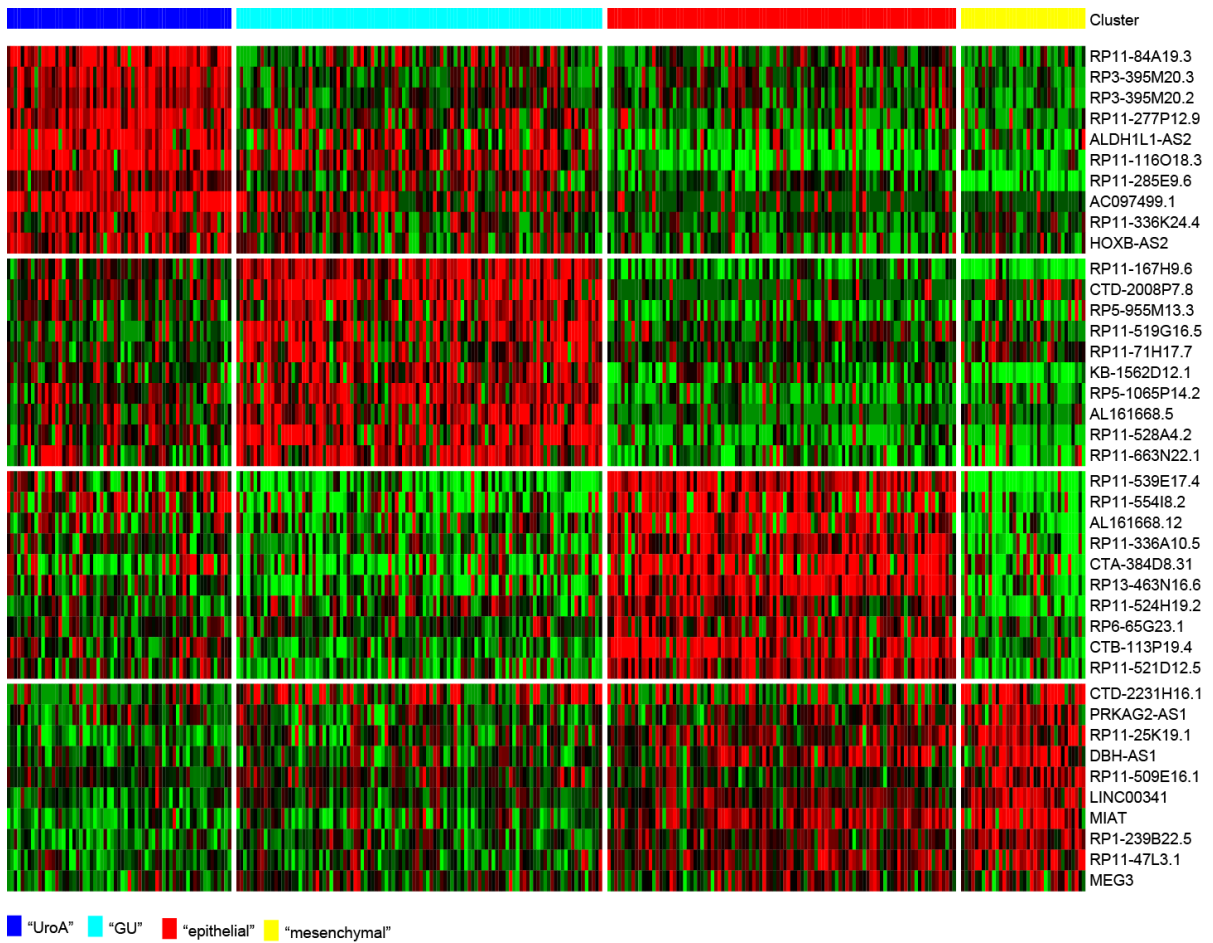
Figure 25: 40-lncRNA signature distinguishes the four lncRNA clusters. The heatmap depicts relative expression of the 40 lncRNAs that distinguish the four lncRNA identified clusters. Color bar: luminal – "UroA" (blue), luminal – "GU" (light blue), basal – "epithelial" (red), basal – "mesenchymal" (yellow).

4.3. Discussion

Emerging evidence implicates long non-coding RNAs (lncRNAs) in the regulation of crucial biological processes in cancer. Using unsupervised hierarchical analyses of whole genome lncRNA expression to identify intrinsic lncRNA subtypes of human muscle-invasive bladder cancers (MIBCs), we identified two lncRNA MIBC subtypes that were almost identical to the recently discovered intrinsic basal and luminal subtypes identified through mRNA expression profiling. The basal lncRNA subtype could be further subdivided into "epithelial' and "mesenchymal" subtypes and the luminal lncRNA subtype could be further subdivided into "UroA" and "GU" subtypes. We see a difference in survival outcomes between the sub-clusters of basal, but there is no survival difference between the sub-clusters of the luminal subtype. However, GSEA did reveal a potential difference in tumor grade between the two luminal sub-clusters.

Analysis of the significant differentially expressed lncRNAs will add a significant layer of complexity to our current understanding of MIBC development and progression. The lncRNAs over-expressed in the basal subtype were commonly associated with squamous cancers or regulators of EMT. Two of the basal lncRNAs, RP11-132A1.4 and RP11-357H14.19, were shown to be up-regulated in hypopharyngeal squamous cell carcinomas and head and neck squamous cell carcinomas, respectively (121, 122). RP11-357H14.19 also plays host to miR-196a, which is also over-expressed in basal MIBCs (122). ZEB1-AS1 is overexpressed in esophageal squamous cell carcinomas, and is also associated with lymph node metastasis and poor overall survival and disease specific survival

(123). Other over-expressed basal lncRNAs include DNM3OS, which is predicted to be under the control of TWIST1, a key regulator of EMT (124). RP11-38P2.2 was also overexpressed in basal MIBCs and is predicted to be controlled by miR-205-5p, a common inhibitor of EMT by targeting ZEB1 (125). However, it is also suggested that this lncRNA acts as a sponge RNA to interact with and suppress miR-205 expression to maintain totipotency of cells (125). The most interesting overexpressed lncRNA in the basal cancers was MEG3, which has been identified to possess tumor suppressive activity as it is lowly expressed in tumor tissues when compared to normal tissues (70, 126-128). In this present study, we see that high expression of MEG3 is associated with a lethal sub-cluster of basal tumors, suggesting that more studies need to be done to determine the role of MEG3 in the progression of MIBC.

Several lncRNAs overexpressed in the luminal subtype were associated with response to cisplatin or with better survival outcomes. RP11-789C1.1 expression was associated with better survival outcomes in gastric cancers, lower rates of metastasis and lower tumor stage (129). RP11-363E7.4 was upregulated after cisplatin exposure, and was associated with up-regulated p53 pathway genes (130). It has been shown that cisplatin treatment upregulates the expression of UCA1, and is further increased in cisplatin-resistant bladder cancer cells (131). Studies have shown that over-expression increases cell viability during cisplatin treatment, whereas UCA1 knockdown partially overcomes drug resistance (131). Since the detection of UCA1 in urine sediment has proven to be highly sensitive and specific for diagnosing bladder carcinoma, it may also be developed into a biomarker to

predict sensitivity to chemotherapy in MIBC (65). UCA1 is also suggested to be regulated by CEBPA by ChIP analyses, a transcription factor found commonly up-regulated in luminal MIBCs (132). However, UCA1 has 3 different splice variants, one of which has been suggested to induce EMT and increase the migratory and invasive abilities of bladder cancer cells by repressing miR-145 expression (133).

Overall, we saw high concordance between the basal and luminal classifications made by using lncRNA and mRNA expression demonstrating that the intrinsic basal and luminal subtypes of MIBC are robust, and that lncRNA expression may serve as a surrogate for identifying mRNA based subtypes. This study also introduces the potential of using relative lncRNA expression to identify tumor subtype. Since the detection of one bladder cancer associated lncRNA in urine sediment has already been established, it should be reasonable to develop a urine based test to quantify lncRNA expression and subsequently identify tumor subtype non-invasively in the near future.

Chapter 5:

Summary & Future Directions

5. Conclusions

Taken together, this dissertation outlines new findings that add significant complexity to the currently understood biology underlying the heterogeneity of MIBC, but also reinforces the presence of intrinsic subtypes of MIBC. This chapter will review the conclusions made from Chapters 3 and 4, and discuss possible future work to expand on the conclusions made.

5.1. Summary

Chapter 3: miRNA expression profiles identify subtypes of MIBC. The work presented in Chapter 3 revolved around two major goals. The first goal was to show that the previously identified intrinsic subtypes of MIBC were reproducible (4, 7, 8, 11). Previous findings were confirmed by subjecting unsupervised analyses to the largest, high-quality data set that is currently available for MIBC from TCGA (n=405). We identified three clusters that were reminiscent of the clusters previously identified (11), including stable basal and luminal subtypes and an "infiltrated" unstable cluster. We then developed an mRNA-PAM classifier that consisted of 593 mRNAs, that was validated in 62 FF samples from our own MDA cohort (GSE48075).

The second goal of this chapter was to identify miRNAs that could optimally distinguish the basal and luminal subtypes of MIBC, which could lead to the development of better diagnostic tools for MIBC. Since there is only one MIBC cohort with matched mRNA and miRNA expression, TCGA's cohort, there was a need to develop a validation cohort. We performed miRNA-sequencing on a cohort

of 62 FF MIBCs, that we have matched whole genome expression profiling data for. An unsupervised approach using miRNA expression alone did not yield positive results, so using a supervised approach we were able to develop a 63-miRNA PAM classifier that identified basal and luminal tumors with more than 80% accuracy. We noticed that the miRNA expression patterns of the basal and luminal subtypes reiterated known biological properties of the subtypes.

Chapter 4: LncRNA expression profiles identify subtypes of MIBC. Since recent work has implicated lncRNAs in the regulation of the hallmarks of cancer, we sought to understand the lncRNA expression patterns that were associated with the basal and luminal subtypes of MIBC. We extracted lncRNA expression data from TCGA's RNA-seq data set (n=407), and performed unsupervised analyses. Based solely on lncRNA expression we identified the basal and luminal subtypes with over 95% accuracy. The lncRNAs overexpressed by basal tumors were associated with transcription factors involved in EMT and associated with squamous cancers, which are both characteristics of basal MIBCs. The overexpressed lncRNAs in luminal tumors were primarily associated with good outcomes, or the development of cisplatin resistance.

We also identified subsets within the basal and luminal subtypes. Within the basal subtype, we identified a larger subset that exhibited "epithelial" properties and a smaller subset that exhibited "mesenchymal" properties, which had poor survival outcomes. The luminal subtype could be further subdivided into a "UroA" subtype,

97

and a "GU" subtype that is enriched with mRNA expression patterns of high grade bladder cancers.

5.2. Future Directions

In order to expand upon the work presented here, there are several follow up experiments that could be performed. First and foremost, the development of a clinical diagnostic tool is greatly needed and should be prioritized. Since ncRNAs are considered to be relatively stable in bodily fluids, and lncRNAs have been previously measured in urine sediment, it should be feasible to develop a clinically relevant tool. However, there are several obstacles that need to be addressed before development would be possible. The first obstacle would be optimizing the RNA isolation protocol. We would need to be able to acquire at least 100 ng of total RNA to perform a custom Nanostring assay with both miRNAs and lncRNAs, or more if miRNA-sequencing and RNA-sequencing is desired. Once the isolation protocol is in place, we would need to decide on either a custom panel of miRNAs and lncRNAs that could identify the intrinsic subtypes or if we want to perform sequencing. Lastly, we would need to perform an independent experiment to confirm that the expression profiles observed in the body fluids are comparable to the expression patterns observed in the primary tumor. The test would then need to be validated and approved for clinical use.

Another important question that needs to be addressed is, which ncRNAs are the most important for basal and luminal gene expression in bladder cancer. This question could be addressed by performing knockdowns of the most deregulated

ncRNAs in basal or luminal cell lines, followed by whole genome expression analysis. If a phenotypic or transcriptomic subtype change were observed after knockdown, then that ncRNA may play an important role in regulating subtype expression profiles. We may also opt to perform similar analyses in breast cancer, to determine which ncRNAs are specific to the basal and luminal subtypes of bladder cancer or to develop a test that could identify basal and luminal subtypes regardless of tumor site. Once these ncRNAs are identified, the molecular mechanism by which they control gene expression would need to be explored. This would be a tremendous undertaking, but it would help the research community understand the underlying biology of basal and luminal expression subtypes. An important distinction that would need to be made is, are the ncRNAs controlling the transcription factors that are implicated as upstream regulators of the subtype, or are the ncRNAs regulating downstream expression of the upstream regulators. This could lead to the subsequent identification of druggable targets, providing MIBC patients with additional treatment options.

Lastly, in order to complete the work presented in chapter 4, it will be important to develop a validation cohort to confirm our findings. This could be addressed by performing RNA-sequencing on our cohort of 62 FF tissues, and subsequently using bioinformatic methods to isolate the lncRNA expression profiles. Ultimately, validating the lncRNA findings in an independent cohort will add a significant amount of relevance to this work.

5.3. Final Discussion

When we look at the data presented as a whole, we see that the expression patterns of ncRNAs can provide insight into the basal and luminal MIBCs since we see similar characteristics of the subtypes regardless of RNA type used to identify subtype. The results also indicate the presence of sub-clusters within the basal and luminal subtypes, which are distinguished by infiltration or the presence of mesenchymal markers. However, a major question that remains is, how exactly do the identified miRNAs and lncRNAs affect the gene expression patterns observed that distinguish the basal and luminal subtypes of MIBC.

One way to begin to address this question is to look at correlations between lncRNA and miRNA expression, as well as the mRNA targets of the miRNA. Using TCGA's cohort, we correlated the overexpressed lncRNAs in luminal tumors with miRNA expression. We observed that several luminal lncRNAs were positively correlated with miR-934, miR-429, miR-200a and miR-200c, and was negatively correlated with miR-146b. When we looked at the lncRNAs overexpressed in basal tumors, we see significant negative correlations with the miR-200 family members (miR-200a, miR-200b, miR-429, miR-141, miR-200c), and positive correlations with miR-142, miR-155, and miR-146b. Based on our previous studies, we see that miR-934, and miR-200 family members are overexpressed in luminal tumors, while miR-146b, miR-142 and miR-155 are overexpressed in basal tumors (114). While the function of miR-934 has not been elucidated, the other miRNAs identified have been implicated in EMT. The miR-200 family directly target transcriptional activators of EMT, including ZEB1/2 (34), miR-155 is positively correlated with metastasis in

breast cancer (88, 134, 135), miR-146b has been implicated in promoting metastasis in some tissues and inhibiting metastasis in others (136, 137), and miR-142 has been implicated as a metastasis suppressor (138, 139).

These results suggest that miRNAs and lncRNAs play a concerted effort to regulate gene expression. MiRNAs have been shown to regulate lncRNAs by binding to regions similar to their mRNA targets causing lncRNA instability, and lncRNAs have been shown to regulate miRNA expression through multiple mechanisms. LncRNAs can regulate miRNA function by acting as a sponge, can directly bind to miRNA to communicate with downstream RNA targets (140), and can compete with miRNA binding to alleviate mRNA suppression (141). LncRNAs can also act as hosts to miRNAs, with approximately 10% of lncRNA genes playing host to miRNAs (142) either in an intron or exon. For example, the H19 lncRNA is a precursor for miR-675 that functions as a tumor suppressor by inhibiting cellular proliferation (143). The interplay between miRNA expression and lncRNA expression implicates ncRNAs as vital modulators of gene expression, however the mechanisms by which they do so remain to be elucidated in MIBC.

Most importantly, this work implicates both lncRNAs and miRNAs as potential diagnostic biomarkers in MIBC. In particular, lncRNA expression analyses were highly consistent with mRNA expression analyses, suggesting an alternative method to diagnosing and characterizing MIBCs. Also, due to their high specificity, lncRNA expression may be a better indicator of disease state.

BIBLIOGRAPHY

1. Society, A. C. 2017. Cancer Facts & FIgures 2017. Atlanta, Ga.

2. Botteman, M. F., C. L. Pashos, A. Redaelli, B. Laskin, and R. Hauser. 2003. The health economics of bladder cancer: a comprehensive review of the published literature. *Pharmacoeconomics* 21: 1315-1330.

3. Shah, J. B., D. J. McConkey, and C. P. Dinney. 2011. New strategies in muscle-invasive bladder cancer: on the road to personalized medicine. *Clin Cancer Res* 17: 2608-2612.

4. Cancer Genome Atlas Research, N. 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507: 315-322.

5. Powles, T., J. P. Eder, G. D. Fine, F. S. Braiteh, Y. Loriot, C. Cruz, J. Bellmunt, H. A. Burris, D. P. Petrylak, S. L. Teng, X. Shen, Z. Boyd, P. S. Hegde, D. S. Chen, and N. J. Vogelzang. 2014. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515: 558-562.

6. Rosenberg, J. E., J. Hoffman-Censits, T. Powles, M. S. van der Heijden, A. V. Balar, A. Necchi, N. Dawson, P. H. O'Donnell, A. Balmanoukian, Y. Loriot, S. Srinivas, M. M. Retz, P. Grivas, R. W. Joseph, M. D. Galsky, M. T. Fleming, D. P. Petrylak, J. L. Perez-Gracia, H. A. Burris, D. Castellano, C. Canil, J. Bellmunt, D. Bajorin, D. Nickles, R. Bourgon, G. M. Frampton, N. Cui, S. Mariathasan, O. Abidoye, G. D. Fine, and R. Dreicer. 2016. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have

progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 387: 1909-1920.

7. Damrauer, J. S., K. A. Hoadley, D. D. Chism, C. Fan, C. J. Tiganelli, S. E. Wobker, J. J. Yeh, M. I. Milowsky, G. Iyer, J. S. Parker, and W. Y. Kim. 2014. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A* 111: 3110-3115.

8. Sjodahl, G., M. Lauss, K. Lovgren, G. Chebil, S. Gudjonsson, S. Veerla, O. Patschan, M. Aine, M. Ferno, M. Ringner, W. Mansson, F. Liedberg, D. Lindgren, and M. Hoglund. 2012. A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res* 18: 3377-3386.

9. Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown, and D. Botstein. 2000. Molecular portraits of human breast tumours. *Nature* 406: 747-752.

10. Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, and A. L. Borresen-Dale. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869-10874.

11. Choi, W., S. Porten, S. Kim, D. Willis, E. R. Plimack, J. Hoffman-Censits, B. Roth, T. Cheng, M. Tran, I. L. Lee, J. Melquist, J. Bondaruk, T. Majewski, S.

Zhang, S. Pretzsch, K. Baggerly, A. Siefker-Radtke, B. Czerniak, C. P. Dinney, and D. J. McConkey. 2014. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 25: 152-165.

12. Amorim, M., S. Salta, R. Henrique, and C. Jeronimo. 2016. Decoding the usefulness of non-coding RNAs as breast cancer markers. *J Transl Med* 14: 265.

13. Liep, J., A. Rabien, and K. Jung. 2012. Feedback networks between microRNAs and epigenetic modifications in urological tumors. *Epigenetics* 7: 315-325.

14. Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75: 843-854.

15. Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 403: 901-906.

16. Pasquinelli, A. E., B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Muller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408: 86-89.

17.   Kaufman, E. J., and E. A. Miska. 2010. The microRNAs of Caenorhabditis elegans. *Semin Cell Dev Biol* 21: 728-737.

18.   Guo, H., N. T. Ingolia, J. S. Weissman, and D. P. Bartel. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466: 835-840.

19.   Eiring, A. M., J. G. Harb, P. Neviani, C. Garton, J. J. Oaks, R. Spizzo, S. Liu, S. Schwind, R. Santhanam, C. J. Hickey, H. Becker, J. C. Chandler, R. Andino, J. Cortes, P. Hokland, C. S. Huettner, R. Bhatia, D. C. Roy, S. A. Liebhaber, M. A. Caligiuri, G. Marcucci, R. Garzon, C. M. Croce, G. A. Calin, and D. Perrotti. 2010. miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell* 140: 652-665.

20.   Calin, G. A., C. D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, and C. M. Croce. 2002. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 99: 15524-15529.

21.   Almeida, M. I., R. M. Reis, and G. A. Calin. 2011. MicroRNA history: discovery, recent applications, and next frontiers. *Mutat Res* 717: 1-8.

22.   Calin, G. A., C. G. Liu, C. Sevignani, M. Ferracin, N. Felli, C. D. Dumitru, M. Shimizu, A. Cimmino, S. Zupo, M. Dono, M. L. Dell'Aquila, H. Alder, L. Rassenti, T. J. Kipps, F. Bullrich, M. Negrini, and C. M. Croce. 2004.

MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* 101: 11755-11760.

23. Croce, C. M. 2009. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 10: 704-714.

24. He, L., J. M. Thomson, M. T. Hemann, E. Hernando-Monge, D. Mu, S. Goodson, S. Powers, C. Cordon-Cardo, S. W. Lowe, G. J. Hannon, and S. M. Hammond. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435: 828-833.

25. Medina, P. P., M. Nolde, and F. J. Slack. 2010. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature* 467: 86-90.

26. Garzon, R., G. A. Calin, and C. M. Croce. 2009. MicroRNAs in Cancer. *Annu Rev Med* 60: 167-179.

27. O'Donnell, K. A., E. A. Wentzel, K. I. Zeller, C. V. Dang, and J. T. Mendell. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435: 839-843.

28. He, L., X. He, S. W. Lowe, and G. J. Hannon. 2007. microRNAs join the p53 network--another piece in the tumour-suppression puzzle. *Nat Rev Cancer* 7: 819-822.

29. Chuang, J. C., and P. A. Jones. 2007. Epigenetics and microRNAs. *Pediatr Res* 61: 24R-29R.

30. Saito, Y., G. Liang, G. Egger, J. M. Friedman, J. C. Chuang, G. A. Coetzee, and P. A. Jones. 2006. Specific activation of microRNA-127 with

downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9: 435-443.

31.   Nicoloso, M. S., R. Spizzo, M. Shimizu, S. Rossi, and G. A. Calin. 2009. MicroRNAs--the micro steering wheel of tumour metastases. *Nat Rev Cancer* 9: 293-302.

32.   Ma, L., J. Teruya-Feldstein, and R. A. Weinberg. 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449: 682-688.

33.   Tavazoie, S. F., C. Alarcon, T. Oskarsson, D. Padua, Q. Wang, P. D. Bos, W. L. Gerald, and J. Massague. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* 451: 147-152.

34.   Gregory, P. A., A. G. Bert, E. L. Paterson, S. C. Barry, A. Tsykin, G. Farshid, M. A. Vadas, Y. Khew-Goodall, and G. J. Goodall. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 10: 593-601.

35.   Gottardo, F., C. G. Liu, M. Ferracin, G. A. Calin, M. Fassan, P. Bassi, C. Sevignani, D. Byrne, M. Negrini, F. Pagano, L. G. Gomella, C. M. Croce, and R. Baffa. 2007. Micro-RNA profiling in kidney and bladder cancers. *Urol Oncol* 25: 387-392.

36.   Catto, J. W., M. F. Abbod, P. J. Wild, D. A. Linkens, C. Pilarsky, I. Rehman, D. J. Rosario, S. Denzinger, M. Burger, R. Stoehr, R. Knuechel, A. Hartmann, and F. C. Hamdy. 2010. The application of artificial intelligence to microarray data: identification of a novel gene signature to identify bladder cancer progression. *Eur Urol* 57: 398-406.

37.   Yates, D. R., I. Rehman, M. F. Abbod, M. Meuth, S. S. Cross, D. A. Linkens, F. C. Hamdy, and J. W. Catto. 2007. Promoter hypermethylation identifies progression risk in bladder cancer. *Clin Cancer Res* 13: 2046-2053.

38.   Catto, J. W., S. Miah, H. C. Owen, H. Bryant, K. Myers, E. Dudziec, S. Larre, M. Milo, I. Rehman, D. J. Rosario, E. Di Martino, M. A. Knowles, M. Meuth, A. L. Harris, and F. C. Hamdy. 2009. Distinct microRNA alterations characterize high- and low-grade bladder cancer. *Cancer Res* 69: 8472-8481.

39.   Dyrskjot, L., M. S. Ostenfeld, J. B. Bramsen, A. N. Silahtaroglu, P. Lamy, R. Ramanathan, N. Fristrup, J. L. Jensen, C. L. Andersen, K. Zieger, S. Kauppinen, B. P. Ulhoi, J. Kjems, M. Borre, and T. F. Orntoft. 2009. Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. *Cancer Res* 69: 4851-4860.

40.   Guancial, E. A., J. Bellmunt, S. Yeh, J. E. Rosenberg, and D. M. Berman. 2014. The evolving understanding of microRNA in bladder cancer. *Urol Oncol* 32: 41 e31-40.

41.   Lowery, A. J., N. Miller, A. Devaney, R. E. McNeill, P. A. Davoren, C. Lemetre, V. Benes, S. Schmidt, J. Blake, G. Ball, and M. J. Kerin. 2009. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res* 11: R27.

42.   Lawrie, C. H., S. Gal, H. M. Dunlop, B. Pushkaran, A. P. Liggins, K. Pulford, A. H. Banham, F. Pezzella, J. Boultwood, J. S. Wainscoat, C. S. Hatton, and A. L. Harris. 2008. Detection of elevated levels of tumour-associated

microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol* 141: 672-675.

43.   Elkin, M., A. Shevelev, E. Schulze, M. Tykocinsky, M. Cooper, I. Ariel, D. Pode, E. Kopf, N. de Groot, and A. Hochberg. 1995. The expression of the imprinted H19 and IGF-2 genes in human bladder carcinoma. *FEBS Lett* 374: 57-61.

44.   Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman. 1990. The product of the H19 gene may function as an RNA. *Mol Cell Biol* 10: 28-36.

45.   Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi, and H. F. Willard. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349: 38-44.

46.   Smith, C. M., and J. A. Steitz. 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol* 18: 6897-6909.

47.   Wang, K. C., and H. Y. Chang. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 43: 904-914.

48.   Lee, C., and N. Kikyo. 2012. Strategies to identify long noncoding RNAs involved in gene regulation. *Cell Biosci* 2: 37.

49.   Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M.

Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigo. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22: 1775-1789.

50.   Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigo, and T. R. Gingeras. 2012. Landscape of transcription in human cells. *Nature* 489: 101-108.

51.   Matouk, I. J., N. DeGroot, S. Mezan, S. Ayesh, R. Abu-lail, A. Hochberg, and E. Galun. 2007. The H19 non-coding RNA is essential for human tumor growth. *PLoS One* 2: e845.

52.   Takai, D., F. A. Gonzales, Y. C. Tsai, M. J. Thayer, and P. A. Jones. 2001. Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum Mol Genet* 10: 2619-2626.

53.   Luo, M., Z. Li, W. Wang, Y. Zeng, Z. Liu, and J. Qiu. 2013. Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett* 333: 213-221.

54.   Ayesh, S., I. Matouk, T. Schneider, P. Ohana, M. Laster, W. Al-Sharef, N. De-Groot, and A. Hochberg. 2002. Possible physiological role of H19 RNA. *Mol Carcinog* 35: 63-74.

55.   Luo, M., Z. Li, W. Wang, Y. Zeng, Z. Liu, and J. Qiu. 2013. Upregulated H19 contributes to bladder cancer cell proliferation by regulating ID2 expression. *FEBS J* 280: 1709-1716.

56.   Qiao, H. P., W. S. Gao, J. X. Huo, and Z. S. Yang. 2013. Long non-coding RNA GAS5 functions as a tumor suppressor in renal cell carcinoma. *Asian Pac J Cancer Prev* 14: 1077-1082.

57.   Kino, T., D. E. Hurt, T. Ichijo, N. Nader, and G. P. Chrousos. 2010. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3: ra8.

58.   Bussemakers, M. J., A. van Bokhoven, G. W. Verhaegh, F. P. Smit, H. F. Karthaus, J. A. Schalken, F. M. Debruyne, N. Ru, and W. B. Isaacs. 1999. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 59: 5975-5979.

59. Crawford, E. D., K. O. Rove, E. J. Trabulsi, J. Qian, K. P. Drewnowska, J. C. Kaminetsky, T. K. Huisman, M. L. Bilowus, S. J. Freedman, W. L. Glover, Jr., and D. G. Bostwick. 2012. Diagnostic performance of PCA3 to detect prostate cancer in men with increased prostate specific antigen: a prospective study of 1,962 cases. *J Urol* 188: 1726-1731.

60. Chen, J., Z. Miao, B. Xue, Y. Shan, G. Weng, and B. Shen. 2016. Long Non-coding RNAs in Urologic Malignancies: Functional Roles and Clinical Translation. *J Cancer* 7: 1842-1855.

61. Barsyte-Lovejoy, D., S. K. Lau, P. C. Boutros, F. Khosravi, I. Jurisica, I. L. Andrulis, M. S. Tsao, and L. Z. Penn. 2006. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* 66: 5330-5337.

62. Ying, L., Q. Chen, Y. Wang, Z. Zhou, Y. Huang, and F. Qiu. 2012. Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol Biosyst* 8: 2289-2294.

63. Han, Y., Y. Liu, Y. Gui, and Z. Cai. 2013. Long intergenic non-coding RNA TUG1 is overexpressed in urothelial carcinoma of the bladder. *J Surg Oncol* 107: 555-559.

64. Yang, C., X. Li, Y. Wang, L. Zhao, and W. Chen. 2012. Long non-coding RNA UCA1 regulated cell cycle distribution via CREB through PI3-K dependent pathway in bladder carcinoma cells. *Gene* 496: 8-16.

65. Wang, X. S., Z. Zhang, H. C. Wang, J. L. Cai, Q. W. Xu, M. Q. Li, Y. C. Chen, X. P. Qian, T. J. Lu, L. Z. Yu, Y. Zhang, D. Q. Xin, Y. Q. Na, and W. F. Chen.

2006. Rapid identification of UCA1 as a very sensitive and specific unique marker for human bladder carcinoma. *Clin Cancer Res* 12: 4851-4858.

66.   Wang, F., X. Li, X. Xie, L. Zhao, and W. Chen. 2008. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS Lett* 582: 1919-1927.

67.   Wang, Y., W. Chen, C. Yang, W. Wu, S. Wu, X. Qin, and X. Li. 2012. Long non-coding RNA UCA1a(CUDR) promotes proliferation and tumorigenesis of bladder cancer. *Int J Oncol* 41: 276-284.

68.   Zhou, Y., Y. Zhong, Y. Wang, X. Zhang, D. L. Batista, R. Gejman, P. J. Ansell, J. Zhao, C. Weng, and A. Klibanski. 2007. Activation of p53 by MEG3 non-coding RNA. *J Biol Chem* 282: 24731-24742.

69.   Zhou, Y., X. Zhang, and A. Klibanski. 2012. MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol* 48: R45-53.

70.   Ying, L., Y. Huang, H. Chen, Y. Wang, L. Xia, Y. Chen, Y. Liu, and F. Qiu. 2013. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. *Mol Biosyst* 9: 407-411.

71.   Srivastava, A. K., P. K. Singh, S. K. Rath, D. Dalela, M. M. Goel, and M. L. Bhatt. 2014. Appraisal of diagnostic ability of UCA1 as a biomarker of carcinoma of the urinary bladder. *Tumour Biol* 35: 11435-11442.

72.   Baldassarre, A., and A. Masotti. 2012. Long non-coding RNAs and p53 regulation. *Int J Mol Sci* 13: 16708-16717.

73. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47.

74. Wilkerson, M. D., and D. N. Hayes. 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26: 1572-1573.

75. McCarthy, D. J., Y. Chen, and G. K. Smyth. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40: 4288-4297.

76. Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.

77. Tibshirani, R., T. Hastie, B. Narasimhan, and G. Chu. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 99: 6567-6572.

78. Lee, W. P., M. P. Stromberg, A. Ward, C. Stewart, E. P. Garrison, and G. T. Marth. 2014. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* 9: e90581.

79. Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome Biol* 11: R106.

80. Lerner, S. P., D. J. McConkey, K. A. Hoadley, K. S. Chan, W. Y. Kim, F. Radvanyi, M. Hoglund, and F. X. Real. 2016. Bladder Cancer Molecular Taxonomy: Summary from a Consensus Meeting. *Bladder Cancer* 2: 37-47.

81.   McConkey, D. J., W. Choi, A. Ochoa, A. Siefker-Radtke, B. Czerniak, and C.
       P. Dinney. 2015. Therapeutic opportunities in the intrinsic subtypes of
       muscle-invasive bladder cancer. *Hematol Oncol Clin North Am* 29: 377-394,
       x-xi.

82.   Mlcochova, H., R. Hezova, M. Stanik, and O. Slaby. 2014. Urine microRNAs
       as potential noninvasive biomarkers in urologic cancers. *Urol Oncol* 32: 41
       e41-49.

83.   Mengual, L., J. J. Lozano, M. Ingelmo-Torres, C. Gazquez, M. J. Ribal, and A.
       Alcaraz. 2013. Using microRNA profiling in urine samples to develop a non-
       invasive test for bladder cancer. *Int J Cancer* 133: 2631-2641.

84.   Shah, M. Y., and G. A. Calin. 2011. MicroRNAs miR-221 and miR-222: a new
       level of regulation in aggressive breast cancer. *Genome Med* 3: 56.

85.   Lin, R. J., D. W. Xiao, L. D. Liao, T. Chen, Z. F. Xie, W. Z. Huang, W. S.
       Wang, T. F. Jiang, B. L. Wu, E. M. Li, and L. Y. Xu. 2012. MiR-142-3p as a
       potential prognostic biomarker for esophageal squamous cell carcinoma. *J
       Surg Oncol* 105: 175-182.

86.   Nassirpour, R., P. P. Mehta, S. M. Baxi, and M. J. Yin. 2013. miR-221
       promotes tumorigenesis in human triple negative breast cancer cells. *PLoS
       One* 8: e62170.

87.   Pinatel, E. M., F. Orso, E. Penna, D. Cimino, A. R. Elia, P. Circosta, P.
       Dentelli, M. F. Brizzi, P. Provero, and D. Taverna. 2014. miR-223 is a
       coordinator of breast cancer progression as revealed by bioinformatics
       predictions. *PLoS One* 9: e84859.

88.	Mattiske, S., R. J. Suetani, P. M. Neilsen, and D. F. Callen. 2012. The oncogenic role of miR-155 in breast cancer. *Cancer Epidemiol Biomarkers Prev* 21: 1236-1243.

89.	Esau, C., X. Kang, E. Peralta, E. Hanson, E. G. Marcusson, L. V. Ravichandran, Y. Sun, S. Koo, R. J. Perera, R. Jain, N. M. Dean, S. M. Freier, C. F. Bennett, B. Lollo, and R. Griffey. 2004. MicroRNA-143 regulates adipocyte differentiation. *J Biol Chem* 279: 52361-52365.

90.	Hamam, D., D. Ali, M. Kassem, A. Aldahmash, and N. M. Alajez. 2015. microRNAs as regulators of adipogenic differentiation of mesenchymal stem cells. *Stem Cells Dev* 24: 417-425.

91.	Guan, X., Y. Gao, J. Zhou, J. Wang, F. Zheng, F. Guo, A. Chang, X. Li, and B. Wang. 2015. miR-223 Regulates Adipogenic and Osteogenic Differentiation of Mesenchymal Stem Cells Through a C/EBPs/miR-223/FGFR2 Regulatory Feedback Loop. *Stem Cells* 33: 1589-1600.

92.	Jiang, H., L. Qu, Y. Wang, J. Cong, W. Wang, and X. Yang. 2014. miR-99a promotes proliferation targeting FGFR3 in human epithelial ovarian cancer cells. *Biomed Pharmacother* 68: 163-169.

93.	Xu, N., P. Brodin, T. Wei, F. Meisgen, L. Eidsmo, N. Nagy, L. Kemeny, M. Stahle, E. Sonkoly, and A. Pivarcsi. 2011. MiR-125b, a microRNA downregulated in psoriasis, modulates keratinocyte proliferation by targeting FGFR2. *J Invest Dermatol* 131: 1521-1529.

94.	Scott, G. K., A. Goga, D. Bhaumik, C. E. Berger, C. S. Sullivan, and C. C. Benz. 2007. Coordinate suppression of ERBB2 and ERBB3 by enforced

expression of micro-RNA miR-125a or miR-125b. *J Biol Chem* 282: 1479-1486.

95.   Dou, C., Y. Wang, C. Li, Z. Liu, Y. Jia, Q. Li, W. Yang, Y. Yao, Q. Liu, and K. Tu. 2015. MicroRNA-212 suppresses tumor growth of human hepatocellular carcinoma by targeting FOXA1. *Oncotarget* 6: 13216-13228.

96.   Stern, D. F. 2000. Tyrosine kinase signalling in breast cancer: ErbB family receptor tyrosine kinases. *Breast Cancer Res* 2: 176-183.

97.   Cantile, M., R. Franco, G. Schiavo, A. Procino, L. Cindolo, G. Botti, and C. Cillo. 2011. The HOX genes network in uro-genital cancers: mechanisms and potential therapeutic implications. *Curr Med Chem* 18: 4872-4884.

98.   Millar, N. L., D. S. Gilchrist, M. Akbar, J. H. Reilly, S. C. Kerr, A. L. Campbell, G. A. Murrell, F. Y. Liew, M. Kurowska-Stolarska, and I. B. McInnes. 2015. MicroRNA29a regulates IL-33-mediated tissue remodelling in tendon disease. *Nat Commun* 6: 6774.

99.   Ogawa, T., M. Iizuka, Y. Sekiya, K. Yoshizato, K. Ikeda, and N. Kawada. 2010. Suppression of type I collagen production by microRNA-29b in cultured human stellate cells. *Biochem Biophys Res Commun* 391: 316-321.

100.   Rebouissou, S., I. Bernard-Pierrot, A. de Reynies, M. L. Lepage, C. Krucker, E. Chapeaublanc, A. Herault, A. Kamoun, A. Caillault, E. Letouze, N. Elarouci, Y. Neuzillet, Y. Denoux, V. Molinie, D. Vordos, A. Laplanche, P. Maille, P. Soyeux, K. Ofualuka, F. Reyal, A. Biton, M. Sibony, X. Paoletti, J. Southgate, S. Benhamou, T. Lebret, Y. Allory, and F. Radvanyi. 2014. EGFR

as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med* 6: 244ra291.

101. Choi, W., B. Czerniak, A. Ochoa, X. Su, A. Siefker-Radtke, C. Dinney, and D. J. McConkey. 2014. Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol* 11: 400-410.

102. Chen, G., J. Peng, W. Zhu, G. Tao, Y. Song, X. Zhou, and W. Wang. 2014. Combined downregulation of microRNA-133a and microRNA-133b predicts chemosensitivity of patients with esophageal squamous cell carcinoma undergoing paclitaxel-based chemotherapy. *Med Oncol* 31: 263.

103. Naito, Y., N. Sakamoto, N. Oue, M. Yashiro, K. Sentani, K. Yanagihara, K. Hirakawa, and W. Yasui. 2014. MicroRNA-143 regulates collagen type III expression in stromal fibroblasts of scirrhous type gastric cancer. *Cancer Sci* 105: 228-235.

104. Akbari Moqadam, F., E. A. Lange-Turenhout, I. M. Aries, R. Pieters, and M. L. den Boer. 2013. MiR-125b, miR-100 and miR-99a co-regulate vincristine resistance in childhood acute lymphoblastic leukemia. *Leuk Res* 37: 1315-1321.

105. McConkey, D. J., W. Choi, Y. Shen, I. L. Lee, S. Porten, S. F. Matin, A. M. Kamat, P. Corn, R. E. Millikan, C. Dinney, B. Czerniak, and A. O. Siefker-Radtke. 2016. A Prognostic Gene Expression Signature in the Molecular Classification of Chemotherapy-naive Urothelial Cancer is Predictive of Clinical Outcomes from Neoadjuvant Chemotherapy: A Phase 2 Trial of

Dose-dense Methotrexate, Vinblastine, Doxorubicin, and Cisplatin with Bevacizumab in Urothelial Cancer. *Eur Urol* 69: 855-862.

106. Nohata, N., T. Hanazawa, H. Enokida, and N. Seki. 2012. microRNA-1/133a and microRNA-206/133b clusters: dysregulation and functional roles in human cancers. *Oncotarget* 3: 9-21.

107. Chen, S., X. Chen, Y. L. Xiu, K. X. Sun, Z. H. Zong, and Y. Zhao. 2014. microRNA 490-3P enhances the drug-resistance of human ovarian cancer cells. *J Ovarian Res* 7: 84.

108. Dadhania, V., M. Zhang, L. Zhang, J. Bondaruk, T. Majewski, A. Siefker-Radtke, C. C. Guo, C. Dinney, D. E. Cogdell, S. Zhang, S. Lee, J. G. Lee, J. N. Weinstein, K. Baggerly, D. McConkey, and B. Czerniak. 2016. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. *EBioMedicine* 12: 105-117.

109. Ponting, C. P., P. L. Oliver, and W. Reik. 2009. Evolution and functions of long noncoding RNAs. *Cell* 136: 629-641.

110. Geisler, S., and J. Coller. 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14: 699-712.

111. Prensner, J. R., and A. M. Chinnaiyan. 2011. The emergence of lncRNAs in cancer biology. *Cancer Discov* 1: 391-407.

112. Tano, K., and N. Akimitsu. 2012. Long non-coding RNAs in cancer progression. *Front Genet* 3: 219.

113. Nana-Sinkam, S. P., and C. M. Croce. 2011. Non-coding RNAs in cancer initiation and progression and as novel biomarkers. *Mol Oncol* 5: 483-491.

114. Ochoa, A. E., W. Choi, X. Su, A. Siefker-Radtke, B. Czerniak, C. Dinney, and D. J. McConkey. 2016. Specific micro-RNA expression patterns distinguish the basal and luminal subtypes of muscle-invasive bladder cancer. *Oncotarget* 7: 80164-80174.

115. Charafe-Jauffret, E., C. Ginestier, F. Monville, P. Finetti, J. Adelaide, N. Cervera, S. Fekairi, L. Xerri, J. Jacquemier, D. Birnbaum, and F. Bertucci. 2006. Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* 25: 2273-2284.

116. Subramanian, M., P. Francis, S. Bilke, X. L. Li, T. Hara, X. Lu, M. F. Jones, R. L. Walker, Y. Zhu, M. Pineda, C. Lee, L. Varanasi, Y. Yang, L. A. Martinez, J. Luo, S. Ambs, S. Sharma, L. M. Wakefield, P. S. Meltzer, and A. Lal. 2015. A mutant p53/let-7i-axis-regulated gene network drives cell migration, invasion and metastasis. *Oncogene* 34: 1094-1104.

117. Cui, X., X. Jing, C. Long, J. Tian, and J. Zhu. 2017. Long noncoding RNA MEG3, a potential novel biomarker to predict the clinical outcome of cancer patients: a meta-analysis. *Oncotarget*.

118. Jaeger, J., D. Koczan, H. J. Thiesen, S. M. Ibrahim, G. Gross, R. Spang, and M. Kunz. 2007. Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clin Cancer Res* 13: 806-815.

119. Jin, C., Y. Zheng, Y. Huang, Y. Liu, L. Jia, and Y. Zhou. 2017. Long non-coding RNA MIAT knockdown promotes osteogenic differentiation of human adipose-derived stem cells. *Cell Biol Int* 41: 33-41.

120. Lindgren, D., F. Liedberg, A. Andersson, G. Chebil, S. Gudjonsson, A. Borg, W. Mansson, T. Fioretos, and M. Hoglund. 2006. Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. *Oncogene* 25: 2685-2696.

121. Zhou, J., W. Li, T. Jin, X. Xiang, M. Li, J. Wang, G. Li, X. Pan, and D. Lei. 2015. Gene microarray analysis of lncRNA and mRNA expression profiles in patients with hypopharyngeal squamous cell carcinoma. *Int J Clin Exp Med* 8: 4862-4882.

122. Darda, L., F. Hakami, R. Morgan, C. Murdoch, D. W. Lambert, and K. D. Hunter. 2015. The role of HOXB9 and miR-196a in head and neck squamous cell carcinoma. *PLoS One* 10: e0122285.

123. Wang, Y. L., Y. Bai, W. J. Yao, L. Guo, and Z. M. Wang. 2015. Expression of long non-coding RNA ZEB1-AS1 in esophageal squamous cell carcinoma and its correlation with tumor progression and patient survival. *Int J Clin Exp Pathol* 8: 11871-11876.

124. Lee, Y. B., I. Bantounas, D. Y. Lee, L. Phylactou, M. A. Caldwell, and J. B. Uney. 2009. Twist-1 regulates the miR-199a/214 cluster during development. *Nucleic Acids Res* 37: 123-128.

125. Xin, W., X. Liu, J. Ding, J. Zhao, Y. Zhou, Q. Wu, and K. Hua. 2015. Long non-coding RNA derived miR-205-5p modulates human endometrial cancer by targeting PTEN. *Am J Transl Res* 7: 2433-2441.

126. Yin, D. D., Z. J. Liu, E. Zhang, R. Kong, Z. H. Zhang, and R. H. Guo. 2015. Decreased expression of long noncoding RNA MEG3 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. *Tumour Biol* 36: 4851-4859.

127. Sun, M., R. Xia, F. Jin, T. Xu, Z. Liu, W. De, and X. Liu. 2014. Downregulated long noncoding RNA MEG3 is associated with poor prognosis and promotes cell proliferation in gastric cancer. *Tumour Biol* 35: 1065-1073.

128. Shen, Y., L. F. Dong, R. M. Zhou, J. Yao, Y. C. Song, H. Yang, Q. Jiang, and B. Yan. 2016. Role of long non-coding RNA MIAT in proliferation, apoptosis and migration of lens epithelial cells: a clinical and in vitro study. *J Cell Mol Med* 20: 537-548.

129. Song, W., Y. Y. Liu, J. J. Peng, H. H. Liang, H. Y. Chen, J. H. Chen, W. L. He, J. B. Xu, S. R. Cai, and Y. L. He. 2016. Identification of differentially expressed signatures of long non-coding RNAs associated with different metastatic potentials in gastric cancer. *J Gastroenterol* 51: 119-129.

130. Wang, P., J. Cui, J. Wen, Y. Guo, L. Zhang, and X. Chen. 2016. Cisplatin induces HepG2 cell cycle arrest through targeting specific long noncoding RNAs and the p53 signaling pathway. *Oncol Lett* 12: 4605-4612.

131. Fan, Y., B. Shen, M. Tan, X. Mu, Y. Qin, F. Zhang, and Y. Liu. 2014. Long non-coding RNA UCA1 increases chemoresistance of bladder cancer cells by regulating Wnt signaling. *FEBS J* 281: 1750-1758.

132. Xue, M., X. Li, W. Wu, S. Zhang, S. Wu, Z. Li, and W. Chen. 2014. Upregulation of long non-coding RNA urothelial carcinoma associated 1 by CCAAT/enhancer binding protein alpha contributes to bladder cancer cell growth and reduced apoptosis. *Oncol Rep* 31: 1993-2000.

133. Xue, M., H. Pang, X. Li, H. Li, J. Pan, and W. Chen. 2016. Long non-coding RNA urothelial cancer-associated 1 promotes bladder cancer cell migration and invasion by way of the hsa-miR-145-ZEB1/2-FSCN1 pathway. *Cancer Sci* 107: 18-27.

134. Baffa, R., M. Fassan, S. Volinia, B. O'Hara, C. G. Liu, J. P. Palazzo, M. Gardiman, M. Rugge, L. G. Gomella, C. M. Croce, and A. Rosenberg. 2009. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *J Pathol* 219: 214-221.

135. Hui, A. B., W. Shi, P. C. Boutros, N. Miller, M. Pintilie, T. Fyles, D. McCready, D. Wong, K. Gerster, L. Waldron, I. Jurisica, L. Z. Penn, and F. F. Liu. 2009. Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Lab Invest* 89: 597-606.

136. Geraldo, M. V., A. S. Yamashita, and E. T. Kimura. 2012. MicroRNA miR-146b-5p regulates signal transduction of TGF-beta by repressing SMAD4 in thyroid cancer. *Oncogene* 31: 1910-1922.

137. Deng, X., B. Wu, K. Xiao, J. Kang, J. Xie, X. Zhang, and Y. Fan. 2015. MiR-146b-5p promotes metastasis and induces epithelial-mesenchymal transition in thyroid cancer by targeting ZNRF3. *Cell Physiol Biochem* 35: 71-82.

138. Schwickert, A., E. Weghake, K. Bruggemann, A. Engbers, B. F. Brinkmann, B. Kemper, J. Seggewiss, C. Stock, K. Ebnet, L. Kiesel, C. Riethmuller, and M. Gotte. 2015. microRNA miR-142-3p Inhibits Breast Cancer Cell Invasiveness by Synchronous Targeting of WASL, Integrin Alpha V, and Additional Cytoskeletal Elements. *PLoS One* 10: e0143993.

139. Lu, Y., N. Ji, W. Wei, W. Sun, X. Gong, and X. Wang. 2017. MiR-142 modulates human pancreatic cancer proliferation and invasion by targeting hypoxia-inducible factor 1 (HIF-1alpha) in the tumor microenvironments. *Biol Open* 6: 252-259.

140. Ebert, M. S., and P. A. Sharp. 2010. Emerging roles for natural microRNA sponges. *Curr Biol* 20: R858-861.

141. Xia, T., Q. Liao, X. Jiang, Y. Shao, B. Xiao, Y. Xi, and J. Guo. 2014. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. *Sci Rep* 4: 6088.

142. Rodriguez, A., S. Griffiths-Jones, J. L. Ashurst, and A. Bradley. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* 14: 1902-1910.

143. Keniry, A., D. Oxley, P. Monnier, M. Kyba, L. Dandolo, G. Smits, and W. Reik. 2012. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol* 14: 659-665.

VITA

Andrea Elizabeth Ochoa was born in McAllen, Texas on June 12, 1990, the daughter of Imelda Ramirez Ochoa and Eli Rene Ochoa. After completing her work at South Texas High School for Health Professions, Mercedes, Texas in 2008, she entered St. Edward's University in Austin, Texas. She received the degree of Bachelor of Science with a major in Bioinformatics and a specialization in Biomathematics from St. Edward's University in May 2011. In August of 2011 she entered The University of Texas MD Anderson Cancer UTHealth Graduate School of Biomedical Sciences.