Spring 5-8-2021

Mucin and Splice Variant Profiles of Pancreatic Adenocarcinoma Predict Patient Survival and Subtyping

Christopher M. Thompson
University of Nebraska Medical Center

Follow this and additional works at: https://digitalcommons.unmc.edu/etd

Part of the Biochemistry Commons, Bioinformatics Commons, Computational Biology Commons, Digestive System Diseases Commons, and the Oncology Commons

Recommended Citation
https://digitalcommons.unmc.edu/etd/541

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@UNMC. It has been accepted for inclusion in Theses & Dissertations by an authorized administrator of DigitalCommons@UNMC. For more information, please contact digitalcommons@unmc.edu.
Mucin and Splice Variant Profiles of Pancreatic Adenocarcinoma Predict Patient Survival and Subtyping

by

Christopher M. Thompson

A DISSERTATION

Presented to the Faculty of
The University of Nebraska Graduate College
in Partial Fulfillment of the Requirement for the
Degree of Doctor of Philosophy

Interdisciplinary Graduate Program in Biomedical Sciences
(Specialization: Biochemistry & Molecular Biology)

Under the Supervision of Professor Surinder K. Batra

University of Nebraska Medical Center
Omaha, NE

April 2021

Supervisory Committee:
Sushil Kumar, Ph.D. Kaustubh Datta, Ph.D.
Maneesh Jain, Ph.D. Joyce Solheim, Ph.D.
Dario Ghersi, M.D., Ph.D.
Acknowledgments

Undertaking a doctoral project demands intense patience, flexibility, determination, curiosity, and openness. However, those characteristics alone are not sufficient to propel candidates to the conclusion of their dissertation research. As such, the acknowledgements might be my favorite section of this thesis: the moment I get to recognize and thank all the people that supported, encouraged, and helped me along the way. The clear starting point is with my mentors: Drs. Surinder Batra and Sushil Kumar. Together, they helped me learn how to focus my curiosity and carefully apply a sense of logic, reasoning, and skepticism to the results and conclusions within my own work and presented in the work of others. Possibly without their exclusive intention, I also learned how resilient I can be. Through the many, many failed experiments, I came to appreciate all the subtle lessons in acquiring negative data. These hard lessons were often accompanied by harder conversations that reinforced my resolve to make the work, well, actually work. Aside from all the skills, acquired foundational knowledge, and ideas, I complete my doctoral training with a newfound sense of purpose and inner strength. The discerning supervisor I received over my training is responsible for the invigorated sense of confidence and tenacity I take forward into my future endeavors. For the many challenges that I overcame during my doctoral education, I am grateful to Dr. Batra and Dr. Kumar for affording me the opportunity to cultivate a capable and competent scientific perspective.

The work presented in this dissertation arose from moments of frustrating confusion and rare lost sights of the bigger objectives, but the joys of discovery
have always had more power. Guiding me along this journey, I have been privileged to work with an outstandingly supportive (and rightly challenging) supervisory committee. My work and successes would never have been possible without the time, patience, and awe-inspiring insight they have provided. I only hope that I have done an adequate job conveying my deep appreciation for their input along the way.

There are many individuals that have enriched my academic and scientific life over the past five years, many more than I have the space to give all due respect and appreciation. I have been surrounded by colleagues who challenged me daily to see different perspectives or approach issues from new angles. Many of these colleagues transcended the thin borders of coworkers and blessed me with the honor of their friendship. I have been blessed to develop wonderful relationships and lifelong friends. I enthusiastically look forward to witnessing each of their accomplishments and successes in our future years.

To save the absolute most important people in my life to last, I will never be able to let my family know how much their love and support has meant to me. My mom and dad have only ever been entirely understanding and encouraging of my work. I am constantly amazed by the intuitive ways they know when to let me vent frustrations and when they know to snap me back to a grounded state. I hope I have made them proud and always will. Without them in my life, I would have never made any of the achievements that I have experienced. Their sacrifices have brought every possibility in my life to reality. I share the success of earning this degree with them.
PDAC is a pancreatic epithelial malignancy and demonstrates aggressive progression and bleak patient prognosis. Despite decades of research, the evolution of novel diagnostics and intervention modalities for PDAC is stagnant. This dissertation explores the characteristic aberrant and elevated expression of mucins in PDAC. Beginning with the hypothesis that mucins are associated with disease aggressiveness, analysis of PDAC patient survival in TCGA revealed no associations between single mucin expression and patient survival. This led to the underlying issue of PDAC tumor cellularity since this disease demonstrates variability in the proportion of cancer cells within the tumor. Tumor purity assessed with the ABSOLUTE computational algorithm is reported for all patient samples in the TCGA PDAC dataset. Using these purity scores, a mathematical correction of epithelial-specific mucin expression was devised. Again, no significant association between PDAC patient survival and mucin expression was found. Therefore, I investigated combinatorial expression of mucins by Spearman’s nonparametric PCA, which resulted in four groups of mutual expression: Group One= MUC7/12/17, Group Two= MUC1/3/13/19/20, Group Three= MUC6/15/22, and Group Four= MUC2/4/5AC/5B/16/21. These four groups were associated significantly with survival outcomes. To determine the biological implications of
these four groups, PCA scores for all patients were correlated to whole transcriptomes. Significantly correlated genes were assessed for biological pathway upregulation. The four pathway composites revealed potential pathological signatures unrelated to previous PDAC classifications, representing novel PDAC subtypes. The role of mucin splice variants (SVs) was assessed and correlated to PDAC patient survival. Bioinformatic studies revealed 12 total mucin SVs significantly associated with survival. Better survival was correlated with expression of four MUC1, one MUC13, and one MUC20 SVs. High expression of two MUC4, one MUC15, one MUC16, one MUC21, and one MUC22 SVs were correlated with worse survival. The correlation between MUC4-sv-215 and MUC13-sv-201 SVs and survival were PCR validated in PDAC patient samples. These MUC4Δ6 prognostic findings contributed to in vitro studies and the development of a novel nanoparticle assay that detects MUC4-sv-215 in patient biofluids. The cumulative impact of the results described here may advance the clinical utility of mucins and associated SVs for improved diagnosis of PDAC.
Table of Contents

Acknowledgments ........................................................................................................ iii
Table of Contents ........................................................................................................ vii
List of Figures ............................................................................................................. xiii
List of Tables ............................................................................................................... xvi
List of Abbreviations ................................................................................................. xvii
List of Equations ........................................................................................................... xix

CHAPTER 1: INTRODUCTION ......................................................................................... 20
  1.A Pancreatic Anatomy .............................................................................................. 21
  1.B Pancreatic Budding and Fusion .......................................................................... 29
  1.C Pancreatic Progenitor Cell Differentiation ...................................................... 30
  1.D Mesenchymal Signaling .................................................................................... 31
  1.E Pancreatic Ductal Adenocarcinoma (PDAC) ................................................... 32
  1.F Pancreatic Diseases and PDAC Development ................................................... 37
    1.F.1. Pancreatic Intraepithelial Neoplasia ............................................................ 37
    1.F.2. Intraductal Papillary Mucinous Neoplasm ............................................... 40
    1.F.3. Pancreatitis ................................................................................................. 41
    1.F.4. Diabetes ...................................................................................................... 46
1.G Alternative Splicing in Cancer ....................................................... 48

1.G.1. Modes of Splicing ........................................................................49

1.G.2. Processivity of the Splicing Reaction ...........................................50


1.G.4. Aberrations in Splicing and Spliceosomal-Associated Proteins....58

1.G.5. Alternative Splicing in Cancer ......................................................60

1.G.6. Possible Advantageous Use of Alternatively Spliced RNAs .........62

1.H Mucins .............................................................................................. 65

1.I Mucins in PDAC ................................................................................ 65

1.J The Cancer Genome Atlas .............................................................. 70

1.K TCGA PDAC Characterization........................................................ 71

1.L PDAC Subtyping .............................................................................. 74

1.L.1. Collisson Subtyping ................................................................. 78

1.L.2. Moffitt Subtyping ................................................................. 79

1.L.3. Bailey Subtyping ................................................................. 87

1.L.4. Law Subtyping ................................................................. 91

1.M Conceptualization of Dissertation Work ....................................... 94
CHAPTER 2: MUCIN EXPRESSION PROFILE IN PANCREATIC DUCTAL ADENOCARCINOMA

2.A Introduction

2.B Results

2.B.1. General outline of patient samples

2.B.2. Mucins expression influenced by malignant cellularity

2.B.3. Mucin-dependent PDAC patient subtyping

2.B.4. Identification of mucin splice variants associated with PDAC survival

2.B.5. Determining PDAC Tumor Cellularity with PCR

2.B.6. Experimental validation of MUC13 and MUC4Δ6 SV

2.C Discussion

2.D Methods

2.D.1. Data Acquisition and Characterization

2.D.2. Realignment of RNA-Seq. Fragments and SV Calling

2.D.3. Sample Cellularity Correction of Mucin Expression

2.D.4. Mucin Splice Variant Validation in Patient Samples
2.D.5. Statistical Analysis ................................................................. 147
2.D.6. Power Analysis ................................................................. 148
2.D.7. Sex as a Biological Variable ........................................... 148

CHAPTER 3: FUNCTIONAL ANALYSIS OF MUC4Δ6 ......................... 150

3.A Introduction .............................................................................. 151

3.B Results ..................................................................................... 154

3.B.1. MUC4WT Knockdown ...................................................... 154
3.B.2. miniMUC4d6 Overexpression .......................................... 155

3.C Discussion .............................................................................. 158

3.D Methods .................................................................................. 162

3.D.2. Lentiviral Infection of Target PDAC Cells ...................... 165
3.D.3. Overexpression of miniMUC4 ........................................... 166
3.D.5. Target Cell Transfection .................................................. 172
3.D.6. Western Blot Validation .................................................... 175
CHAPTER 4: NANOPARTICLE-BASED CANCER-SPECIFIC ALTERNATIVE SPLICING VARIANTS DETECTION ASSAY

4.A Introduction ................................................................................... 178

4.B Results ........................................................................................... 186

4.B.1. AuNP Synthesis ......................................................................... 186

4.B.2. AuNP-Probe Conjugation Quantification ................................. 189

4.B.3. Detection of MUC4 Wild-type vs MUC4Δ6 Synthetic Target ....... 192

4.B.4. Detection of KRAS Mutations ..................................................... 199

4.B.5. Sample Fragmentation ............................................................... 206

4.B.6. Detection in Murine Serum ......................................................... 206

4.B.7. Serum Preparation for Assay Detection .................................... 211

4.C Discussion ..................................................................................... 214

4.D Methods and Materials ................................................................. 219

4.D.1. Synthesis of AuNPs ................................................................. 219

4.D.2. Functionalization of NPs .......................................................... 222

4.D.3. Particle Quality Assessment and Quantification ...................... 224
List of Figures

Figure 1.1: Pancreatic Anatomy ................................................................. 23
Figure 1.2: Endocrine and Exocrine Cells of the Pancreas ...................... 26
Figure 1.3: Histological Architecture of the Pancreas ............................. 28
Figure 1.4: Pancreatic Ductal Adenocarcinoma ...................................... 35
Figure 1.5: Histology of PanIN Lesions .................................................... 39
Figure 1.6: Histological Features of Pancreatitis ...................................... 44
Figure 1.7: Modes of Alternative Splicing ............................................... 52
Figure 1.8: Intermediates of RNA Splicing .............................................. 54
Figure 1.9: Two Principal Splicing Sites .................................................. 57
Figure 1.10: Structural Motifs of Mucins ............................................... 67
Figure 1.11: Genomic Profiling of PDAC Tumors from TCGA ................. 73
Figure 1.12: Assessment of PDAC Tumor Purity ...................................... 76
Figure 1.13: Collisson Subtyping of PDAC Tumors ................................. 81
Figure 1.14: Moffitt PDAC Subtyping Establishes Two Tumor and Two Stromal Types ................................................................. 84
Figure 1.15: Combining Moffitt Tumor and Stromal Types Demonstrates Survival Differences .............................................................. 86
Figure 1.16: Bailey PDAC Profiling Establishes Four Subtypes ............... 90
Figure 1.17: Law et al. Subtyping Derives Four New PDAC Subtypes with Overlap of Previous Subtyping Studies ................................. 93
Figure 2.1: TCGA PDAC Patient Demographics ...................................... 104
Figure 2.2: Cellular Composition of TCGA PDAC Samples ..................... 107
Figure 2.3: Expression of Mucins in the TCGA PDAC Dataset ....................... 109

Figure 2.4: Expression of Mucins in the ICGC PDAC Comparison Dataset and Survival Associations ................................................................. 112

Figure 2.5: Cellularity-Correct Expression of Mucins in TCGA Generates Four Novel Clusters ................................................................................. 116

Figure 2.6: PCA Clusters of Mucins in TCGA and ICGC Overlap ............... 119

Figure 2.7: Mucin Signatures Demarcate Unique Pathway Activations .... 122

Figure 2.8: PCA-Significant Mucins Independently Predict Association to MUC-Based Subtypes .................................................................................. 125

Figure 2.9: Mucin Splice Variants Correlate with Differential Survival Outcome in PDAC Patients .............................................................................. 127

Figure 2.10: Genes Panel for PCR-based Cellularity Assessment .......... 132

Figure 2.11: Validation of the expression of MUC13-sv205 and MUC4-sv215 .................................................................................................................... 134

Figure 3.1: Domain Moiety of MUC4 .............................................................. 153

Figure 3.2: MUC4Δ6 Expression Enrichment by shRNA Knockdown of Wild-Type ........................................................................................................... 157

Figure 3.3: MiniMUC4Δ6 Overexpression PCR in COLO357 Cells .......... 160

Figure 3.4: pSecTag C-miniMUC4Δ6 Engineering ........................................ 168

Figure 3.5: Restriction Digestion and Ligation of pST-miniMUC4Δ6 ......... 174

Figure 4.1: Synthesis of RNA-Specific Gold NanoProbes ...................... 181

Figure 4.2: Photoswitchable Fluorescent Probes are Detectable After Enzymatic Cleavage ...................................................................................... 183
Figure 4.3: Activation of FAM-Probe Requires Perfect Hybridization of Prey RNA to Bait Sequence .............................................................. 185

Figure 4.4: Verification of Gold Nanoseeds by DLS ........................................ 188

Figure 4.5: Physical Characteristics of AuNProbes ........................................ 191

Figure 4.6: FAM Standard Curve Facilitates Calculation of DNA Probes per AuNP .......................................................................................... 195

Figure 4.7: Limit of Detection of MUC4Δ6 RNA with AuNProbe-MUC4Δ6 . 198

Figure 4.8: AuNProbe-MUC4Δ6 Demonstrates A High Specificity to MUC4Δ6 Target ......................................................................................... 201

Figure 4.9: MUC4WT RNA Does Not Interfere with MUC4Δ6 Detection ..... 203

Figure 4.10: Validation of KRAS AuNProbe Specificity ................................. 205

Figure 4.11: COLO357 RNA Fragmentation Validation ................................. 208

Figure 4.12: Detection of MUC4Δ6 in Orthotopic Plasma .............................. 210

Figure 4.13: Modification of Orthotopic Plasma Preparation Using Proteinase K Pretreatment ................................................................. 213

Figure 4.14: EDTA-Free Proteinase K Treatment of Orthotopic Plasma .... 216
List of Tables

Table 1: Median Mucin Expression Between Stratified TCGA PDAC and ICGC Samples. .......................................................... 114

Table 2: Mucin Primers .......................................................................................................................... 143

Table 3: Housekeeping Gene GPI Primers .................................................................................. 144

Table 4: Cellularity Primers ........................................................................................................ 146

Table 5: shRNA Against MUC4WT .......................................................................................... 163

Table 6: pSecTag C-miniMUC4Δ6 Primers ............................................................................. 170

Table 7: AuNP Concentration Calculations ........................................................................... 193

Table 8: Synthetic Target RNA Sequences .................................................................. 196

Table 9: AuNP Probe Sequences ......................................................................................... 223
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSOLUTE</td>
<td>ABSOLUTE Purity Algorithm</td>
</tr>
<tr>
<td>ADEX</td>
<td>Aberrantly Differentiated Endocrine Exocrine (PDAC Subtype)</td>
</tr>
<tr>
<td>ADM</td>
<td>Acinar-Ductal Metaplasia</td>
</tr>
<tr>
<td>ALS</td>
<td>Amyotrophic Lateral Sclerosis</td>
</tr>
<tr>
<td>AMOP</td>
<td>Adhesion-associated domain in MUC4 and Other Proteins</td>
</tr>
<tr>
<td>AMY2A</td>
<td>Amylase 2A</td>
</tr>
<tr>
<td>AP</td>
<td>Acute Pancreatitis</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>AuNP</td>
<td>Gold Nanoparticle</td>
</tr>
<tr>
<td>AuNPProbe</td>
<td>Gold NanoProbe</td>
</tr>
<tr>
<td>BC</td>
<td>Breast Cancer</td>
</tr>
<tr>
<td>BCR-ABL</td>
<td>Breakpoint Cluster Region protein - tyrosine-protein kinase ABL1 fusion gene</td>
</tr>
<tr>
<td>BRCA1/2</td>
<td>BReast CAncer gene 1/2</td>
</tr>
<tr>
<td>CA19-9</td>
<td>Carbohydrate Antigen 19-9 (aka Sialyl-Lewis Antigen)</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CIP</td>
<td>Calf Intestinal Phosphatase</td>
</tr>
<tr>
<td>CK19</td>
<td>Keratin 19</td>
</tr>
<tr>
<td>CNA</td>
<td>Copy Number Analysis</td>
</tr>
<tr>
<td>CP</td>
<td>Chronic Pancreatitis</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic Light Scattering</td>
</tr>
<tr>
<td>DMD</td>
<td>Duchenne Muscular Dystrophy</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DNTP</td>
<td>Deoxynucleotide Triphosphate</td>
</tr>
<tr>
<td>DSN</td>
<td>Double-Stranded Nuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>EGF</td>
<td>Epithelial Growth Factor</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-Mesenchymal Transition</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen Receptor</td>
</tr>
<tr>
<td>ESRP2</td>
<td>Epithelial Splicing Regulatory Protein 2</td>
</tr>
<tr>
<td>FAM</td>
<td>6-Carboxyfluorescein</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast Growth Factor</td>
</tr>
<tr>
<td>FOLFIRINOX</td>
<td>Combination Chemotherapy [Leucovorin, Fluorouracil, Irinotecan, Oxaliplatin]</td>
</tr>
<tr>
<td>FPKM</td>
<td>Fragments per Kilobase of transcript per Million mapped reads</td>
</tr>
<tr>
<td>GDC</td>
<td>Genomic Data Commons</td>
</tr>
<tr>
<td>GPI</td>
<td>Glucose-6-Phosphate Isomerase</td>
</tr>
<tr>
<td>ICGC</td>
<td>International Cancer Genome Consortium</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>IPMN</td>
<td>Intraductal Papillary Mucinous Neoplasm</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani (broth or agar)</td>
</tr>
<tr>
<td>MAGEH1</td>
<td>Melanoma-associated AntiGen H1</td>
</tr>
<tr>
<td>MHC</td>
<td>Major Histocompatibility Complex</td>
</tr>
<tr>
<td>MODY</td>
<td>Maturity-Onset Diabetes of the Young</td>
</tr>
<tr>
<td>MSP-RON</td>
<td>Macrophage Stimulating Protein-receptor tyrosine kinase Recepteur d'Origine Nantais</td>
</tr>
<tr>
<td>MUC</td>
<td>Mucin (followed by specific gene number)</td>
</tr>
<tr>
<td>NIDO</td>
<td>Nidogen-like Domain</td>
</tr>
<tr>
<td>NP</td>
<td>Nanoparticle</td>
</tr>
<tr>
<td>PAAD</td>
<td>TCGA Pancreatic Cancer Dataset</td>
</tr>
<tr>
<td>PACA</td>
<td>ICGC Australian Pancreatic Cancer Dataset</td>
</tr>
<tr>
<td>PC</td>
<td>Pancreatic Cancer</td>
</tr>
<tr>
<td>PC1/2/3/4</td>
<td>Principal Component 1/2/3/4</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDAC</td>
<td>Pancreatic Ductal Adenocarcinoma</td>
</tr>
<tr>
<td>PDI</td>
<td>Polydispersity Index</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>PR</td>
<td>Progesterone Receptor</td>
</tr>
<tr>
<td>PRPF</td>
<td>Pre-mRNA Processing Factor genes</td>
</tr>
<tr>
<td>PSA</td>
<td>Prostate Specific Antigen</td>
</tr>
<tr>
<td>PST</td>
<td>pSecTag-C Plasmid</td>
</tr>
<tr>
<td>PTK6</td>
<td>Protein Tyrosine Kinase 6</td>
</tr>
<tr>
<td>QM</td>
<td>Quasimesenchymal (PDAC Subtype)</td>
</tr>
<tr>
<td>RFU</td>
<td>Relative Fluorescent Units</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>SEER</td>
<td>Surveillance, Epidemiology, and End Results Program</td>
</tr>
<tr>
<td>SF3B1</td>
<td>Splicing Factor 3 Binding Protein 1</td>
</tr>
<tr>
<td>SMCC</td>
<td>Succinimidyl Cyclohexane Maleimide</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SPINK1</td>
<td>Serine Peptidase Inhibitor Kazal Type 1</td>
</tr>
<tr>
<td>SV(s)</td>
<td>Splice Variant(s)</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-Carboxyethyl)Phosphine</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>TCR</td>
<td>T-Cell Receptor</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission Electron Microscopy</td>
</tr>
<tr>
<td>TIDM</td>
<td>Type I Diabetes</td>
</tr>
<tr>
<td>TIIDM</td>
<td>Type II Diabetes</td>
</tr>
<tr>
<td>TPM</td>
<td>Transcripts per Million reads mapped</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>WES</td>
<td>Whole Exon Sequencing</td>
</tr>
</tbody>
</table>
List of Equations

Equation 1: FPKM to TPM Conversion .............................................................. 141
Equation 2: Calculation for PDAC Sample Cellularity by PCR ...................... 145
Equation 3: Beer-Lambert Equation ................................................................. 186
Equation 4: Calculation of Probes per AuNP ................................................. 192
CHAPTER 1:

INTRODUCTION
The pancreas is the largest glandular organ in the human body, and its functions and development are intriguingly complex. Unique among all human glands, the pancreas has both endocrine and exocrine secretions and provides vital roles in digestion and homeostasis. While much has been discovered about the physical course of pancreatic embryology, the specific molecular signaling pathways involved in the regulation of its tissue differentiation have proven to be more elusive and shrouded in convolution.

Furthermore, disorders and diseases of the pancreas present with life-altering consequences and their respective treatments, albeit often failing to lead to cures in patients, range from long-term maintenance to short-term, palliative intervention. Also, medical treatment of the pancreas bridges many different specialties from endocrinology, to gastrointestinal, to immunology, and at times surgery, with each frequently having opposing approaches to therapy.

I begin with an overview of the anatomy of the pancreas and explore the development of the human pancreas starting at the first collections of cells at the 26th day of embryonic development and follow it through the multilayers of the cell and inter-organ tissue signaling. I will then proceed into three of the most common diseases of the pancreas and attempt to explore what contributes to the complexity of their establishment and treatments.

1. A Pancreatic Anatomy

Retroperitoneally and transversely along the posterior abdominal wall, posterior to the stomach, in the left upper abdomen, the pancreas is a large organ utilizing considerable space. Structurally, the pancreas is comprised of three main
Figure 1.1: Pancreatic Anatomy

The pancreas runs midline-left, laterally across the upper abdomen inferior to the liver and diaphragm and posterior to the stomach.

Used with permission, © Terese Winslow, 2009. (Appendix A-1)
FIGURE 1.1: PANCREATIC ANATOMY
sections: the head, which is posterior and dextrolateral to the stomach, the body, oriented mediodorsal to the stomach, and the tail, situated sinistrolateral to the stomach and ventral to the left kidney. (Fig 1.1) Inferior to the liver, the head of the pancreas receives the common bile duct superiorly and presents the anastomosed bile duct and pancreatic duct to the duodenal lumen via the major duodenal papilla of Vater under smooth muscular control of the sphincter of hepatopancreatic ampulla. In most people, the vestigial proximal portion of the dorsal pre-pancreatic bud, known as the accessory pancreatic duct of Santorini, remains intact and fuses to the duodenum superior to the major papilla.

The major functions of the pancreas can be divided into two aspects (Fig 1.2). The digestive roles are performed by exocrine cells composed of acinar cells, which produce and secrete enzymes, and ductal cells, which conduct those enzymes from the acini to the central pancreatic duct then to the duodenal lumen where they mix with chyme. The hormonal functions are conducted by clusters of endocrine cells arranged in cords juxtaposed to acini and ducts known as islets of Langerhans. Pancreatic islets are composed of glucagon-producing α-cells, insulin-producing β-cells, pancreatic polypeptide-producing γ-cells, somatostatin-producing δ-cells, and ghrelin-producing ε-cells (Fig 1.3). Together, the endocrine population comprises about 2% of the total mass of the pancreas whereas the exocrine population makes up the remaining 98%. The establishment of these two very distinct populations of cells from a common progenitor origin is complex, requiring many cooperative and competitive signaling cascades.
Figure 1.2: Endocrine and Exocrine Cells of the Pancreas

The human pancreas is located left shifted across the midline in the upper abdomen and is comprised of endocrine and exocrine cell populations, giving the pancreas dual biological functions.

Used with Permission, cancer.org (Appendix A-2)
FIGURE 1.2: ENDOCRINE AND EXOCRINE CELLS OF THE PANCREAS

- Endocrine cells secrete hormones into blood vessels
- Exocrine cells secrete pancreatic enzymes into the pancreatic duct
- Pancreas
- Pancreatic duct
- Liver
- Gallbladder
- Common bile duct
- Ampulla of Vater
- Duodenum
- Duct to pancreatic duct
Figure 1.3: Histological Architecture of the Pancreas

The functional histology is comprised of exocrine cells retained in acini secreting digestive products into ductal lumen and clusters of endocrine cells releasing their hormonal and peptidyl products into vascular circulation.

Used with permission from 'Exocrine Pancreas' by Ed Friedlander MD, Department of Pathology, Kansas City University, Kansas City, MO (Appendix A-3)²
FIGURE 1.3: HISTOLOGICAL ARCHITECTURE OF THE PANCREAS
1.B Pancreatic Budding and Fusion

In the first trimester of embryonic development, 26 days after gestation, a small cluster of cells can be observed on the posterior aspect of the gut tube, juxtaposed to the duodenal anlage distal to the stomach, yielding the first evidence of pancreatic development. Following the expansion of these cells and preceding loss of contact between the gut tube and notochord due to the fusion of the two dorsal aortas, the endoderm invades the overlaying mesoderm. These initial differentiation regulatory signals have been suggested to come from the notochord-derived morphogens via sonic hedgehog (Shh) inhibition in the pancreatic precursor cells. The absence of Shh in the pancreas stands in stark contrast to the rest of the gastrointestinal system where it remains an essential pathway to organogenesis. Six days following this event, a second bud arises from the caudal hepatic bud representing the ventral pancreatic bud. Shortly before the presentation of the ventral pancreatic bud, the epithelium of the coelom rapidly divides and sequesters the gut and pancreas from the aorta and non-gut tissues while mesenchyme simultaneously proliferates and begins to separate the pancreatic and coelomic epithelium. The epithelium of both pancreatic ducts begins to proliferate and elongate in a stalk-like manner with branching points arising at acute angles in a fashion that omits intervening mesenchyme between branches. Thus, this morphological development results in a tissue comprised of yet an indistinguishable cell population with no distinctive subcellular structures. Between embryonic day 37 and 42, the gut tube experiences a posterio-anterior loop elongation and physiological hernia into the umbilicus followed by clockwise
rotation in which the pylorus shifts dextrolateral as the duodenum twists establishing its final C-loop. As a consequence of this rotation, the ventral and dorsal pancreatic buds come into contact, and the fusion of the buds and their ducts ensue.

1. C Pancreatic Progenitor Cell Differentiation

To this point, all pancreatic endocrine cells have appeared primarily glucagon positive. However, following the initiation of gut rotation, the pancreas enters a phase known as the "secondary transition" in which endocrine cell populations increase in dramatic fashion and shift toward insulin-secreting β-cells\(^6\) and enzyme granules appear in dense concentration in newly differentiated acinar cells distinguishable from ductal cells by their expression of pancreatic specific transcription factor \(1\alpha\) (PTF1\(1\alpha\)).\(^7\) Distal epithelial cells begin expressing PTF1\(1\alpha\) around E12 and subsequently form a spherical lumen. Initiating between E14 and E15, these PTF1\(1\alpha\)-positive cells increase in size, shift their nuclei to the basal aspect, and demonstrate eosinophilic cytoplasm as digestive secretions are increasingly upregulated.\(^8\) Specifically in insulin-positive endocrine cells, formation of cords of endocrine-positive cells begins as the cells down-regulate adhesion molecules, shift their plane of proliferation by 90 degrees, and begin migrating away from the basement membrane of the epithelial proto-lumen.\(^4\) It has been suggested that this transition is also marked by a reduction in pancreatic duodenal box-1 (Pdx1) expression which has been present in the developing pancreatic cells since the generation of the pre-pancreatic buds in a process likely analogous to epithelial-to-mesenchymal transition.\(^9, 10\) However, Gannon and colleagues used
a Pdx1-knockout mouse model and demonstrated that the mice failed to establish entire populations of endocrine-positive islets.\textsuperscript{11}

1.D Mesenchymal Signaling

Early studies of the differentiation of the pancreatic epithelium into endocrine islets and exocrine acini revealed the presence and selective contact of pancreatic mesenchyme to pancreatic epithelia. In the complete absence of mesenchyme, pancreatic epithelium collectively differentiated to endocrine islets.\textsuperscript{12} When mesenchyme was present and contacting pancreatic epithelium, acinar differentiation dominated.\textsuperscript{13} The mediator of this effect was discovered to be the induction of Notch, enhancement of downstream hairy enhancer of split-1 (Hes1), and consequential neurogenic-3 (Ngn3) inhibition.\textsuperscript{14} Given the lack of co-localized expression of E-cadherin and Ngn3\textsuperscript{15}, procession toward terminal pancreatic differentiation is likely due to an initial, temporal inhibition of islet development following inhibitory signaling of Ngn3 and enhanced Hes1 in the endocrine-negative epithelium. More illuminating, Shih and group demonstrated that the differentiation signaling might be dose-dependent, in which high Notch was able to activate Hes1 and SRY-Box Transcription Factor 9 (Sox9) resulting in suppression of Ngn3 and retention of a ductal phenotype.\textsuperscript{16} Conversely, low Notch signaling was able to activate Sox9 but failed to activate Hes1. Thus Ngn3 was expressed, and endocrine differentiation was established. In a somewhat contradictory manner, it has also been observed that mesenchymal presence is capable of expanding Pdx-1 positive cell populations, likely required to generate
sufficient numbers of endocrine progenitor cells preceding endocrine terminal differentiation.\textsuperscript{17}

A number of molecules have been identified in pancreatic mesenchyme-induced signaling including fibroblast growth factors (FGF), which have well-documented histories in mesenchymal interactions with epithelium, suggested to play critical roles in mesenchymal-derived notch signaling and subsequent Ptf1a expression \textsuperscript{18}, enhance amylase expression \textsuperscript{17}, and cause significant suppression of endocrine cell populations \textsuperscript{19}. Although, interestingly, FGF7 enhancement has been observed to increase epithelial growth and result in reduced endocrine differentiation.\textsuperscript{20}

\textbf{1.E Pancreatic Ductal Adenocarcinoma (PDAC)}

Pancreatic cancer (PC) is one of the deadliest malignancies, currently on track to become the second most lethal by 2020. The 2018 projections for PC from the latest Surveillance, Epidemiology, and End Results Program (SEER) estimate that over 55,000 new cases will be diagnosed and over 44,000 deaths will occur.\textsuperscript{21} Despite considerable advances made in diagnostics and therapeutics in the field of oncology, the goals of early detection of and novel chemotherapeutics for PC are unmet. With over half of all PC patients diagnosed with distant metastatic lesions, the 5-year survival rate has remained an abysmal \textasciitenvisioned{\sim}10\%. Other dismal, innate characteristics of PC contributing to its substantial lethality while denigrating the efficiency of therapeutic intervention include low immune cell infiltration\textsuperscript{22, 23}, early metastasis\textsuperscript{24}, inherent drug resistance\textsuperscript{25-27}, and high desmoplasia\textsuperscript{25, 26}. Therefore, advances in earlier diagnosis are likely to be most beneficial and
temporally opportune for PC patients.\textsuperscript{28} Surgical resection remains the most successful PC intervention; however, many patients fail to meet the criteria for surgery making chemoradiotherapy their sole option.\textsuperscript{26, 28, 29} Still, of the patients that do qualify for surgical resection, a subset demonstrates poor response and ultimately succumb to either post-surgical complications or more rapid disease progression for reasons still unknown.\textsuperscript{30} Aberrant early onset and progressive expression of many species of mucins are characteristic of PC.\textsuperscript{31-35} Within PDAC, the most prevalent and lethal PC, the tumor mass contains a unique variable and relatively low population of malignant cells. Much of the bulk of PDAC tumors is comprised of extracellular components, collagens, and other fibrotic proteins (\textbf{Fig 1.4}).\textsuperscript{36} The fibrosis has been documented by surgical teams to be so dense that a core necrotic cyst sometimes develops.\textsuperscript{36} Furthermore, many pathways have been demonstrated to be aberrantly altered and may contribute to the high aggressiveness and lethality of PDAC. Interestingly, many of these dysregulated pathways are represented in the initial section of this discussion on the embryological establishment of the pancreas. Many of the histological observations associated with pre-lesions of PDAC are similar to the differentiation processes during embryological development and comprise convergent pathways to PDAC. Pancreatic intraepithelial neoplasm, one of the most recognizable malignant transformations during PDAC initiation is marked by a degenerative loss of the ordinarily well-defined apical border of the ductal lumen into a pseudo-brush border-like, undulating epithelium. Intraductal papillary mucinous neoplasm (IPMN) are also common initiating conditions for PDAC and possess ductal cells
Figure 1.4: Pancreatic Ductal Adenocarcinoma

Infiltrating ductal carcinoma of the pancreas presents with characteristic disruption of cuboidal epithelium morphology and the abundance of desmoplastic stroma. Green dashes in 4x image mark adenocarcinoma invasion, blue dash marks adjacent PanIN3 lesion. Black arrows of 10x image indicate invasive carcinoma. Brown arrow of 20x image indicates the invasive low-grade carcinoma from the duct.
FIGURE 1.4: PANCREATIC DUCTAL ADENOCARCINOMA
Chapter I: Introduction

with abundant, pale staining cytoplasm containing abundant glycoprotein products known as mucins. Several mucins (MUC) have become well documented in PDAC with aberrant and increasing expression. Of exception, MUC1, MUC4, MUC5AC, and MUC16 have individually been shown to be drastically elevated in a progression-dependent manner in PDAC cases and may correlate with differential aggressiveness of tumors. \(^{37}\) Much work is being conducted regarding the use of mucins as biomarkers.

Arguably one of the most stimulating topics related to PDAC pathology may be the concept of acinar-to-ductal metaplasia (ADM). Similar to epithelial re-differentiation to mesenchymal cells of the stromal compartment during wound healing, acinar cells can undergo a reprogramming following pancreatic tissue insult, but this process is observed to be dysregulated and rapidly occurring in PDAC. \(^{38}\) Histological presentation of ADM is marked by alteration of the nuclear shape into a thinner, more flat appearance. This process may represent a reversal of one of the last stages of functional differentiation in the embryonic pancreas where the pre-pancreatic cells receive a poised-ductal programming signal from the notochord initially and the mesenchyme subsequently. The ultimate consequence of these compartment alterations is that the appreciable architecture of the pancreas becomes highly disordered with acini, ducts, and islets lose their defining characteristics. The most well recognized molecular findings in PDAC are expression of constitutively activated KRAS, loss of Smad4 and p16, and mutated p53. \(^{39}\) While countless other mutations and aberrant signaling pathways have
been described, these four persist and often are shown to play roles in “novel” mutation pathways discovered in vitro and in vivo PDAC studies.

1.F Pancreatic Diseases and PDAC Development

The route from tissue insult to pancreatic cancer is an obfuscated process. While the link between noncancerous and cancerous conditions of the pancreas have been postulated and superficially evaluated, the complete picture of the actual process is yet to be entirely elucidated. These disorders may well contribute to the pathogenesis of PDAC and require additional investigations to determine their potential roles in disease biology. As such, these possible involvements necessitate an introduction to the pathobiology of some common pancreatic conditions.

1.F.1. Pancreatic Intraepithelial Neoplasia

Pancreatic intraepithelial neoplasms (PanINs) are histologically defined precursor lesions marked by a progressive disorientation of ductal architecture (Fig 1.5).\textsuperscript{40} Due to the histological identification of PanIn lesions, their discovery often occurs serendipitously in tissue adjacent to resected PDAC tumors. While literature suggests that PanINs progress into PDAC, likely due to their proximity to and the histological similarities to PDAC tumors, a lack of consensus regarding their implication on PDAC aggressiveness or time-to-tumorigenesis is still grossly lacking.\textsuperscript{41-43} Interestingly, studies have demonstrated that cells within acinar-ductal metaplastic (ADM) lesions, characterized by dedifferentiation of acinar cells observed after pancreatic tissue insult, acquire the commonly PDAC-associated
Figure 1.5: Histology of PanIN Lesions

A normal pancreatic duct loses its cuboidal morphology in the first step of transformation to pancreatic intra-epithelial neoplasia. This initial change is designated PanIN1A, which progresses to PanIN1B when the cells lose polarity. Low-grade PanIN lesions, classified as benign, progress with loss of basement membrane (PanIN2) and eventually high-grade PanIN3 with nuclear apolarity, hyperchromatic nuclei, macronucleoli, and luminal atrophy (black dash focal point of lesion). In the PDAC Image, high-grade PanINs (blue dash) frequently associated with invasive carcinoma (black dash).
FIGURE 1.5: HISTOLOGY OF PANIN LESIONS
activating KRAS mutations only in ADMs associated with PanINs.\textsuperscript{44} This driver mutation acquisition has historically been the attributable factor in defining PanINs as PDAC precursor. However, new evidence from PanINs and PDAC tumors resected from patients suggests that the progression to malignancy is not a direct stepwise progression, but may occur via at least three different pathways.\textsuperscript{45} In one scenario, both PDAC and PanIN cells have distinct mutational signatures indicating that the two lesions arose independently from a common ancestral cell. The second scenario demonstrated that the same mutations were present in both PanIN cells and PDAC tumor cells; however, the latter acquired additional mutations that may have contributed to their malignant transformation. In the third scenario, the mutational signatures in PDAC and PanIN cells were identical, implying the mutational burden gave rise to a forward progressive transformation. Still, understanding the genomic alterations that occur within PanINs and how they relate to the changes that occur in pancreatic adenocarcinoma cells may distinguish yet unknown biomarkers and/or therapeutic targets for improved clinical approaches.

1.F.2. Intraductal Papillary Mucinous Neoplasm

Intraductal papillary mucinous neoplasia (IPMN) is a second histologically defined precursor lesion of the pancreas in which tumors develop within the ductal lumen. Approximately 10\% of IPMNs progress to invasive PDAC.\textsuperscript{46} These lesions are also remarkable in that they produce large quantities of mucins. Acutely, IPMNs may contribute to symptoms associated with occluded pancreatic ducts and are frequently appreciable after they progress to mucin-filled ductal cysts.
Previously termed ‘Mucinous Pancreatic Adenomas’, IPMN lesions are now classified along a broad spectrum from low-grade mild dysplasia all the way up to severe invasive carcinoma.\textsuperscript{47} The radiographic and histological appearances of IPMNs can be idiosyncratic to their origins and location, but the robust expression of MUC5AC and punctate expression of MUC1 are typical\textsuperscript{48}. When assessing the mutational burden of IPMNs, 50\% of patient lesions were found to carry KRAS codon 12 mutations while mutations in the Go protein subunit was detected in 79\%.\textsuperscript{49} The presence of these mutations may indicate a possible mechanism by which IPMNs transform into invasive malignancy.

1.F.3. Pancreatitis

Pancreatitis is a multi-faceted disease of variable etiology with elaborate clinical and laboratory presentations and patient-specific symptomology, aside from nearly universal upper quadrant pain. However, histological examination of pancreatitis samples reveals massive fibrosis and collagen fiber deposition, alongside fat and parenchymal necrosis visible by gross pathological examination.\textsuperscript{50} The underlying mechanism contributing to the development of pancreatitis is thought to be the premature activation of zymogens or digestive enzymes.\textsuperscript{50}

About the several sub-classifications of pancreatitis and specialized diagnostic approaches to each one, indistinguishable alterations in stromal composition, fibrosis, and scarring (in the most advanced cases) persist as the universal histological characteristics of all subtypes. Of note, however, autoimmune pancreatitis demonstrates a unique case of histological distinction in
that the infiltration of leukocytes contributes to an appreciable difference in the tissue appearance. Further, depending on the mode of immune infiltration, additional dissimilarities become apparent with inflammation contributing to fibrotic alterations of ductal basement membranes or the generation of fibrotic lobules. As in the case of all other forms of pancreatitis, histological clues inferring the stage of progression may be apparent when considering the cell density and degree of fibrosis at presentation (Fig 1.6)\(^51\). The ultimate and most notable presentation of fibrosis is the generation of vacuole-like gaps in stroma as the collagen is crosslinked and becomes progressively more irregular and denser. These gaps may appear at any location throughout the pancreas from the surface and deeper following the deposition of collagen.

The clinical presentation of pancreatitis is varied but most often includes intense pain in the middle upper abdomen and digestive malabsorption. The most intense pain coincides with the historically labeled acute pancreatitis (AP) and is associated with a near 10% mortality \(^52\). An interesting area of focus is the transition from AP to chronic pancreatitis (CP), which occurs in approximately 7% of AP cases with 11-15% of AP patients presenting with a recurrent AP attack \(^53,\) \(^54\). Currently there is a transition to contract the understanding of AP and CP as distinct conditions but rather ranges of severity of the same diagnosis. Still, the clinical presentations, risks, treatments, and outcomes of AP are grossly different from CP.\(^55\) As such, the two terms now have distinct use in clinical management with restricted use in pancreatitis research. AP patients have rapidly increased pancreatic enzyme premature activation coinciding with pancreatic cell necrosis.
The progression of pancreatitis produces histology that transitions with disease. Cell-rich, storiform fibrosis observable during early pathology. Large focal areas of architectural disruption are hallmarks of sustain pancreatitis (black arrows). With chronic disease, a prolonged immune infiltration and inflammation (red circle of insert) become notable. The protracted inflammation results in intense fibrotic scaring and consist primarily of collagen fibrils and reduced cell populations. Blue arrow of insert image indicates PanIN2 with possible concurrent IPMN.
Chapter I: Introduction

FIGURE 1.6: HISTOLOGICAL FEATURES OF PANCREATITIS
The condition induces a strong systemic inflammatory reaction and, at its most severe, leads to multiorgan failure and death. CP typically occurs with less severe pathology but over many years compared to days with AP. The prolonged tissue insult in CP results in slow loss of pancreatic function, contributing to poor digestive abilities, nutritional deficiencies, and diabetes. Additionally, the rate of pancreatic cell loss or turnover most frequently associates with pancreatic fibrosis compared to necrosis in AP. Therapeutic approaches to AP usually involve pain management and fluid support, whereas intermittent fasting, lifestyle changes, and partial pancreatic resection can all be indicated treatment for CP.

Substantial research has investigated the risk factors associated with pancreatitis. Nearly 80% of acute attacks of pancreatitis diagnoses are related to alcohol use or the presence of gallstones. While other physical risks contributing to pancreatitis have been identified, such as drug use and obesity, reported statistical assessment of these risks varies among studies and across geographical regions. However, emerging evidence has suggested that the use of smoking tobacco may impart a risk of pancreatitis greater than that observed with binge alcohol abuse. Finally, several genetic variations have been linked to lifelong risk of pancreatitis. Extrapancreatic factors, like genomic variations in alcohol dehydrogenase or aldehyde dehydrogenase genes, are associated with altered ability to metabolize alcohol effectively, while other variants in pancreatic associated genes, such as trypsinogen and Serine Peptidase Inhibitor Kazal Type 1 (SPINK1) among others, may constitute premature enzyme activation or reduced/inhibited secretion of digestive enzymes.
1.F.4. Diabetes

Diabetes mellitus is a chronic condition marked by glucose uptake dysregulation affecting approximately 8.5% of the global adult population, according to the World Health Organization. Historically identified as a hormonal disorder, research has demonstrated that diabetes is more a systemic disorder affecting everything from neural health, to vascular function, to immune response, and tissue healing processes. While some minor forms exist, diabetes is predominantly classified into two major categories: Type I and Type II.

Type I diabetes mellitus (TIDM) is most easily explained by the immune-mediated destruction of the pancreatic insulin-producing beta cells (β-cells) in the islets of Langerhans. Importantly, inflammation has been suggested as the most influential contributor to β-cell degradation. Due to lack of an available detectable biomarker, knowledge of the initiation and progression of the disease is severely incomplete. Clinicians do know, however, that destruction of pancreatic β-cells follows an observable infiltration of lymphocytes into islets and upregulation of insulin-positive β-cell-associated human leukocyte antigen (HLA-ABC). This presentation, known as insulinitis, has been documented in up to 23% of insulin-positive islets in 78% of recent-onset type 1 diabetes patients. Following the elevation of HLA expression in islets, detection of insulin-positive endocrine cells is drastically reduced. Consequentially, patients experience a profound, life-altering symptomology associated with type 1 diabetes. Systemic insulin levels are critically depleted, and maintenance of normal-ranged serum glucose levels depends exclusively on routine post-prandial administration of pharmaceutical
recombinant insulin, frequent blood glucose checks, and often the use of glucose supplements to rescue dangerous drops in serum glucose because of impaired glucagon function.

Type II diabetes mellitus (TIIIDM) may easily be viewed as an entirely separate disease than TIDM, in that pancreatic β-cells are present within islets and produce insulin during the initial phase and but quickly deplete as systemic resistance to insulin signaling drives β-cells to exhaustion and islet inflammation increase.\textsuperscript{62} This increased inflammation is one of the histological hallmarks of a TIIIDM pancreas and has been shown to be stimulated, at least in part, by increased production of interleukin-1β (IL-1β) from pancreatic islet β-cells exposed to prolonged, high levels of serum glucose.\textsuperscript{24} This abundant IL-1β mediates two distinct responses: one, β-cells experience a subsequent upregulation in NF-KB transcriptional activity and Fas-elicited apoptosis; and two, islet-resident macrophages expressing receptors for IL-1β induce inflammasome formation, contributing to and sustaining an inflammatory environment. Furthermore, secretions of islet amyloid polypeptide (IAPP; pre-cleaved peptide co-secreted with insulin) from β-cells have been hypothesized to contribute substantially to amyloid plaque deposition in diabetic islets, even when insulin-deficient β-cells persist in the pancreatic islets.\textsuperscript{63}

The connection between diabetes and PDAC is not straightforward; however, evidence suggests a calculable risk exists between the two. Meta-analysis has shown that longstanding diabetes coincides with a nearly 2-fold increase in PDAC development risks over an individual’s lifetime, while newly
diagnosed diabetes elevates the risk to as high as 8-fold.\textsuperscript{64} The calculated risk of PDAC development with new-onset diabetes is now used as a clinical predictor indicating higher PDAC screening in older individuals known as the ENDPAC model.\textsuperscript{65} Independent validation of ENDPAC scoring further identified concurrent weight loss as a risk factor of clinical importance raising the baseline prevalence of PDAC in control subjects from 0.78\% to 1.7\% in comorbid patients.\textsuperscript{66} While the determinants between PDAC and diabetes overlap, the underlying pathways contributing to both conditions are important areas for future research.

\textbf{1.G Alternative Splicing in Cancer}

Modern concepts of cancer have focused on the concepts of precursor conditions, acquisition of mutational burden, uncontrolled proliferation, desmoplasia, immune modulation, and altered cellular metabolism. However, developments in technologies and increasing computational power have advanced our ability to begin to decipher the complex cell transcriptomes. Next-generational sequencing has revealed a staggering array of RNA splice variants in all malignancies, but the significance of their functions or pathological relevance are poorly understood for the most part. At present, up to 95\% of the 20,000-25,000 human genes are believed to undergo alternative splicing.\textsuperscript{67} While the mis-splicing of critical genes might seem to be an insulting cellular event, it may actually be beneficial to life by providing the templates for a substantial number of potentially advantageous novel proteins. Alternative splicing has been attributed to physiological homeostasis within cells, most famously with the splicing of Bcl-x of the Bcl-2 family with the shorter isoform (Bcl-xS) contributing to apoptosis and the
longer isoform (Bcl-xL) having counter, antiapoptotic functions.\(^68\) Still, a higher prevalence of alternative splicing in cancers has been observed many times.\(^69\text{-}72\) This suggests that cancers capitalize on the random accumulations of genetic lesions and aberrant splicing which contribute to proteomic neofunctionalism to promote survival and aggressiveness.\(^73\text{-}80\) Additionally, tumor cells may use these novel proteins for other aspects, such as regulatory decoys, metastasis, immune evasion, or therapeutic resistance. This further leads to the question of potentially targeting these new alternative transcripts clinically for diagnosis, prognostic consideration, or treatment. While a few notable alternatively spliced genes have been well studied regarding their function and specific expression in cancers, the majority remain obscure. While the field of cancer-specific splicing continues to develop, several considerations require attention, including altered mechanisms of splicing (i.e., splice site mutations, mutations in spliceosomal components, altered expression of spliceosome-associate proteins, or preference for U12-minor splicing), cancer-type specific expression of alternate splices, and functionality of novel splices. Here, I will review the most recent studies investigating the structure of the spliceosome, mechanisms of splicing, and cancer-specific alternatively spliced genes with the intention of identifying potential splice variants of interest for diagnosis, prognosis, or therapeutic targeting.

1.G.1. Modes of Splicing

Splicing of the pre-messenger RNA (pre-mRNA) is a well-documented and vital process of cell biology. There are around 30,000 known genes coded in the \textit{Homo sapiens} genome representing roughly only 1% of the prodigious 3 billion
Chapter I: Introduction

base pairs organized into 23 pairs of chromosomes. The remaining 99% of the genome is comprised of intragenic non-coding sequences and intergenic introns that separate the exons of the final protein coding sequences. While these intronic sequences present a seemingly dilemmatic problem for the study of genomic sequences of proteins, it is well recognized that inclusion of these additional bases provides adaptive potential to organisms from an evolutionary perspective. Several types of splicing scenarios to account for the inclusion introns or exclusion of exons have been reported (Fig 1.7).^{81}

1.G.2. Processivity of the Splicing Reaction

To understand the errors of splicing and the associated consequences, a review of the spliceosome, or splicing machinery complex, and the reactions it facilitates are necessary. The mechanisms of RNA splicing have been well investigated in vitro and processing ‘decisions’ by the cell to use the major or minor splicing machinery are determined by the exon-adjacent intronic sequences, though the general reactions overlap between the two pathways requiring 5 small nuclear RNAs (snRNAs) for both splicing classifications which interact with as many as 170 nuclear proteins^{82} to form small nuclear ribonucleoproteins (snRNPs). The snRNAs facilitate RNA-RNA and RNA-protein interactions while the proteins of the snRNPs are responsible for the stabilization of the spliceosome complexes and catalytic processes of splicing. In a regulated, stepwise set of interactions and reactions, four intermediate complexes are formed before the final spliced exon product and the cleaved intron are released (Fig 1.8).^{83} The general process of intronic splicing, undergone concurrently with transcription, is
**Figure 1.7: Modes of Splicing**

Schematic diagram depicting the various methods of splicing events that potentially exist in cells. Under normal splicing regulation, introns are removed, and all coded exons are ligated to form the final sequence of the mature messenger RNA. Some introns may be included in the final sequence due to mis-splicing, consequently altering the structure and potentially ameliorating the translatability of a (dys)functional protein. Mutual exclusion of exons seems to be an evolutionarily retained mechanism which allows a single gene to code for multiple functional proteins. Under mutual exclusion, the two affected exons are never retained in the final mRNA together. Exon skipping is the exact opposite of intron inclusion whereby an entire exon is removed within the lariat loop. Finally, the presence of cryptic splice sites can alter the pivot points of the spliceosome and alter the final coding sequence of the resulting mRNA.
FIGURE 1.7: MODES OF ALTERNATIVE SPLICING

Normal Splicing

Intron Inclusion

Mutual Exclusion of Exons

Exon Skipping

Alternative Splice Sites
Splicing of pre-mRNA generates 4 intermediate species. Recognition of the 5’ splice site by U1 initiates splicing and generates Complex E. The recruitment of U2/U2AF upstream of the 3’ splice site sequence establishes Complex A and primes the pre-mRNA for structural rearrangement. Binding of the tri-snRNP U4/U5.U6 folds the RNA creating the lariat loop of Complex B. Release of U1 and U4 brings the catalytic domains of the spliceosome to the RNA creating Complex C. Cleavage and ligation of the ribose backbone fuse the adjacent exons and the spliceosome disassembles.
FIGURE 1.8: INTERMEDIATES OF RNA SPlicing

pre-mRNA

\[ \text{Exon 1} \quad \text{GU} \quad \text{A} \quad \text{AG} \quad \text{Exon 2} \]

Intron

Complex E

\[ \text{Exon 1} \quad \text{GU} \quad \text{A} \quad \text{AG} \quad \text{Exon 2} \]

\[ U1 \]

Complex A

\[ \text{Exon 1} \quad \text{GU} \quad \text{A} \quad \text{AG} \quad \text{Exon 2} \]

\[ U2 \]

\[ U4 \]

\[ U6 \]

Complex B

\[ \text{Exon 1} \quad \text{GU} \quad \text{A} \quad \text{AG} \quad \text{Exon 2} \]

\[ U4 \]

\[ U7 \]

Complex C

\[ \text{Exon 1} \quad \text{GU} \quad \text{A} \quad \text{AG} \quad \text{Exon 2} \]

Splice mRNA

\[ \text{Exon 1} \quad \text{Exon 2} \]
mechanistically the same and involves the recognition of the 5' splice site sequence, invariable adenine within the branch point, and the 3' splice site sequence (Fig 1.9). Following the formation of the lariat loop, two esterification reactions release the intron and bond the retained exons. While the mechanisms of major and minor splicing are the same, the initiation of splicing is different between the two.


The major spliceosome, as its name suggests, is the most frequent method of RNA splicing. The 5' splice site is recognized and bound by snRNP U1 forming the first intermediate Complex E. The splicing factor U2 Auxiliary Factor (U2AF) complex binds the polypyrimidine sequence upstream of the 3' splice site. In an ATP-dependent manner, the snRNP U2 binds the pre-mRNA just upstream of the U2AF binding site to form the second intermediate Complex A. Next, the U4/U5.U6 complex engages the spliceosome with U4 unwinding and U6 undergoing a conformational transformation upon interacting with the 5' end of the intron. U6 binds U2 at the 3’ end forming a lariat loop while U5 binds the junctional ends of the adjacent exons establishing Complex B. U1 and U4 are displaced to form Complex C as the spliceosome prepares to perform two esterification reactions at the adjacent exonic sequences. The exons are ligated as the lariat intron is released and the splicing machinery disassembles.

A small population of introns demonstrate sequences at the 5' (AT) and 3' (AC) splice sites that deviate from the recognized and bound sequences described above and entirely absent of the polypyrimidine sequence. It was found that the
Figure 1.9: Two Principal Splicing Sites

The consensus sequences of the 5’ and 3’ splice sites differ between two predominate pre-mRNA classes. The contained sequence determines if the RNA is spliced via the major or minor spliceosome pathway by the recruitment and binding of spliceosomal-specific snRNPs and associated binding proteins. The blue boxes highlight the major spliceosomal sequences while red boxes indicate minor spliceosomal sequences. Internal black boxes indicate the consensus binding sites. The order (top to bottom) and size of each base letter indicates the relative proportion at that site.
FIGURE 1.9: TWO PRINCIPAL SPLICING SITES

5' Splice Site

Branch Point

3' Splice Site

Major Spliceosome Seq  Minor Spliceosome Seq
processing spliceosomes of these introns contained the low frequency U11 and U12 instead of U1 and U2/U2AF. In this minor splicing reaction, U11 dissociates from U12 to bind at the 5’ site while U12 binds the 3’ site forming Complex E and Complex A, respectively. The lariat loop of Complex B is formed when the tri-snRNP U4atac/U5.U6atac dissembles allowing U6atac to binds the 5’ site, undergo a dynamic transformation, and bind U12 while U5 stabilizes the junctional sequences of both exons. U11 is displaced forming Complex C. The same two esterification reactions cleave the intron and ligate the exons releasing the lariat intron before the spliceosome complex disassembles.

1.G.4. Aberrations in Splicing and Spliceosomal-Associated Proteins

While splicing is strongly credited with providing cells with crucial adaptive abilities and superior gene expression regulation during embryologic development, it is also heavily involved in disease processes. Alternative splicing is generally accepted to be the consequence of several different errors. The first group of possible causes are mutations of the key splicing regulatory domains called cis-acting elements and include the core consensus sequences (5’ and 3’ splice sites and the branch point) and splicing enhancers or silencers located in the intronic or exonic sequences. These mutations can be within the genomic sequence or a read/write error in the transcription of messenger RNA (mRNA). Several pathogenic conditions have been associated with mutations in cis-element sequences leading to mis-splicing of the disease-related mRNA. One of the most described involves mutation in the dystrophin gene, expressed mostly in muscular and skeletal cells, and causes a family of X-linked conditions called Duchenne and
Becker muscular dystrophy. Dystrophin is expressed in three functional isoforms based on the presence of independent promoters. \(^{85}\) However, point mutations in the dystrophin gene contribute to many alternative truncated or frameshifted isoforms of final mRNA contributing to the progressive loss of neuromuscular function. Around 25% of dystrophin mutations involve changes in splice site sequences. \(^{86}\) The mutation HBB110G>A contributes to the creation of an incorrect 3' splice site and causes the very well-known disease β-thalassemia. \(^{87}\) Many mutations in cis-elements of the intermediate filament lamin A/C gene LMNA lead to mis-splicing of the several LMNA mRNA expressed from the single locus and have been linked to upwards of 14 different diseases ranging from muscular dystrophies to cardiomyopathies. \(^{88-92}\) In many of these aforementioned diseases, gene therapies are being investigated to modify the spliced mRNA products and have shown great promise for potentially successful treatment of afflicted patients.

The second classification of aberrant splicing is caused by mutations within the core spliceosome-associated proteins. While it is intuitive that mutations in splicing associated proteins may have horribly negative consequences, the number of proteins involved in spliceosomal complexing and function in combination with incomplete penetrance makes predicting outcomes of such mutations a difficult achievement. Diseases associated with spliceosome mutations are more rare owing to the vital nature of its function. However, mutations in pre-mRNA processing factor genes (PRPF) have been documented to interfere with U4/U5.U6 tri-snRNP complex formation, causing several conditions involving mis-spliced genes. One of the more completely described of
these mutations is found in PRPF6, leading to the inhibition of U4/U6 interaction in retinitis pigmentosa.\textsuperscript{93}

The third class of splicing-affected mutations involve those in trans-acting splicing factors. Owing to the involvement of splicing factors in the recruitment of the core spliceosome to pre-mRNA, alterations in sequence or concentrations of extrinsic trans factors can have complex consequences on the final mRNA sequences. More directly, these consequences involve inclusion of introns or exclusion of exons constituting changes in functional domains of the translated protein. Because splicing factors’ functions are intrinsically linked to their structural motifs, modifications in trans-acting factors can fail to recruit spliceosome snRNPs to either splice site by mutation-induced loss of protein interactions or can mask the splice sites altogether. An example of conditions caused by this type of dysregulation is the infamous amyotrophic lateral sclerosis, more specifically, familial hereditary ALS. Chromosomal expansion of a G4C2 sequence in the C9orf72 has been suggested to sequester splicing factors contributing to a global dysregulation of pre-mRNA splicing.\textsuperscript{94} Still, other ALS-associated mutations of splicing trans factors can have damaging consequences on the splicing of other spliceosome proteins, such as the heterogenous nuclear ribonucleoprotein P2 coded by the FUS gene. FUS is involved in the splicing of numerous genes, one of which is U1.\textsuperscript{95}

1.G.5. Alternative Splicing in Cancer

Relevant to the direction of this dissertation, alternative splicing is heavily involved in cancer biology. RNA-sequencing studies have linked the abundance of
alternative splicing events in cancer cells to the abnormal elevation of premature stop codons found in RNA of cancers compared to normal cells. Indeed, a recent expansive pan-cancer study demonstrated that cancers possess nearly 20% more alternatively spliced RNA than normal cells. The discovery that many point mutations in cancers are linked to loss of splice site sequence combined with the more recent finding that cancer-specific mutations generate novel splice sites leads to the question of why these cells are not targeted by immune surveillance cells for destruction. Interestingly, this same study found that many of these splicing affected mutations were found in immune checkpoint genes including those involved in PD1-PDL1 pathways. However, other examples of the previous classes of splicing dysregulation are also abundant in cancers. Sequencing analysis of hematopoietic cells revealed mutations in the U2AF1 gene and contributed to an accumulation of mis-spliced mRNA in cells of several hematologic malignancies. While the presence of these mutations in cancers are of interest, more relevant is the contribution these genomic alterations make to the pathologies of their respective disease states, of which many groups have written very succinct analyses. Some more notable examples involve many aberrant alternative splice sites in the master cell cycle regulator Tp53 consequently reducing the anti-proliferative response of p53, bimodal alteration in apoptosis by alternative caspase 2 and caspase 8, enhanced cell mobility by altered interaction of macrophage-stimulating protein receptor and Recepteur d’Origine Nantais, and promotion of angiogenesis by mis-splicing vascular endothelial growth factor A. The splicing dysregulation having
direct consequences on malignant processes are described as most interesting because these cancer-specific sequences lend themselves to the development of novel diagnostic detection methods and targeted therapeutic strategies, as discussed in the next section.

1.G.6. Possible Advantageous Use of Alternatively Spliced RNAs

While detection of alternative splicing *per se* is unhelpful, the identification of cancer-specific alternative splice variants as biomarkers has the potential to tremendously progress our ability to diagnose malignancies in patients early and potentially increase the likelihood of a successful intervention. Mutational analyses of many genes have highlighted involvement in diseases, and now a growing body of evidence is emerging on splice variants, whether from genomic mutation or afflicted splicing regulation. The tumor suppressors BRCA1 and BRCA2 are frequently implicated in many malignancies, including cancer of the breast, ovary, and pancreas. A Breast Information Core database analysis of BRCA1 and BRCA2 demonstrated that nearly 5% of their alterations in breast cancer were due to alternative splice site utilization. Importantly, however, the expression of alternative splice variants is found throughout the body in many tissues. Therefore, the specificity of a particular variant in tissues of interest must be weighed against its expression in non-target tissues for its use as a relevant biomarker. For example, an investigation of the variant BRCA2Δ12 was found expressed 33% higher in hormone-negative breast tumors compared to normal breast tissue. The discovery of prostate-specific antigen (PSA), encoded by the kallikrein-3 (KLK3) gene, has revolutionized our clinical ability to screen and monitor men for
the development and progression of prostate cancer. However, other conditions, such as prostatitis or benign prostate hyperplasia, can cause detectable elevations in PSA. Alternatively, studies looking at KLK genes expressed in prostate cancer have found that low expression of the prostate-specific KLK11 correlated with a more aggressive disease.\textsuperscript{114}

Detection of splice variants for diagnostic purposes would be a massive clinical advancement; however, detection of these variants plausibly benefits those diagnosed in early disease. Therefore, the therapeutic targeting of cancer-specific splice variants stands to benefit those unlucky to receive advanced stage diagnoses. Employment of alternatively spliced RNA targeting will require extensive consideration and study as the consequences of many are unknown or under investigated. Methods of capitalizing on alternative splicing may involve the design of small molecules to counteract the effects of biologically active variant proteins or targeting and neutralizing the variant mRNA itself.\textsuperscript{80} One very promising example of the latter is the use of splice-editing antisense oligonucleotides in Duchenne muscular dystrophy. DMD is a complex disease containing few deletion events resulting in a truncated dystrophin protein. A very common mutation in dystrophin causes a frameshift mutation in exon 51. Studies have used a few antisense oligonucleotide molecules (drisapersen, eteplirsen, and golodirsen) to mask and induce subsequent skipping of exons 51 and 53, contributing to a truncated but functional dystrophin protein.\textsuperscript{115, 116} These therapeutic techniques require substantially more study but have shown promise in early clinical trials. An emerging concept in cancer intervention is the role splice variants may play in
therapeutic resistance in tumor cells. The presence of a point mutation in the BCR-ABL fusion protein (caused by chromosomal rearrangement found in chronic myeloid leukemia and other cancers) contributes to the alternatively spliced BCR-ABL35Ins.\textsuperscript{117} This variant has been linked to poor response to the first-line therapy imatinib, a tyrosine kinase inhibitor, due to a loss of a drug-binding site in the variant’s modified conformational shift.\textsuperscript{118-120} The loss of function of BRCA1 and BRCA2 in tumor cells forces a reliance on poly-ADP-ribose polymerase (PARP) to repair DNA strand breaks for genomic integrity. Some clinical trials are exploring the use of PARP inhibitors as chemotherapeutic sensitizers in BRCA1/2 deficient cancers with some success.\textsuperscript{121-123} Other strategies to directly correct mis-splicing are also hypothesized to have significant benefits. A mutation in the BRAF gene (BRAFV600E) is highly implicated in the aggressiveness of the skin cancer melanoma, present in around 60% of patient tumors.\textsuperscript{124} Treatment of BRAFV600E-expressing melanomas frequently involves BRAF inhibitors. However, some tumor cells express a truncated variant that skips exons 4-8. This variant is not bound by BRAF inhibitors, due to modification of its ATP-binding site. The use of spliceostatin A analogues has been suggested to target the trans-acting factor SF3B1 thereby inhibiting the formation of the core spliceosome snRNP U2-SF3B1 complex.\textsuperscript{125} This approach has demonstrated the ability to prevent exon skipping, therefore recovering BRAF inhibitor sensitivity.\textsuperscript{80, 126} With each study published, strategies to correct the inbuilt advantages that alternative splicing may provide to cancer cells evolves, but we still have much left to uncover.
1.H Mucins

Mucins are large glycoproteins containing compulsory tandem repeat domains with complex carbohydrate branches enzymatically added through a series of post-translational reactions (Fig 1.10). Class-wise, mucins are sorted into two distinct groups: transmembrane or secreted. Physiological expression of mucin is apically observed on epithelial cells, especially in the respiratory mucosa and gastrointestinal tract. The functions of mucin include establishing a physical barrier and mediating cell-cell and cell-stroma interactions, cell signaling, and cell migration.

1.I Mucins in PDAC

The role of mucins in PC disease biology has been reported extensively, including their contributions to cell migration and metastasis, proliferation, and drug resistance. With recent advances in high-throughput transcriptomic sequencing and bioinformatics techniques, we now know that cancer cells demonstrate elevated levels of alternative splicing. It is unclear whether spliceosomes act aberrantly, are guided to key splicing sights, or become overwhelmed by transcriptional machinery. However, the expression and detection of disease-specific isoforms hold incredible diagnostic and therapeutic promise. Furthermore, mucins demonstrate numerous alternatively spliced isoforms which can alter the many functional domains coded in their proteins. The Cancer Genome Atlas (TCGA) is a significant resource to assess cancer patient-derived tumor transcriptomes. Through TCGA,
Figure 1.10: Structural Motifs of Mucins

(Top) Mucin structure demonstrates the large architecture of membrane-tethered mucins and their extensive glycan branches. (Bottom) Secreted mucins contain similar structural motifs as transmembrane mucins but lacking the transmembrane domain.

FIGURE 1.10: STRUCTURAL MOTIFS OF MUCINS

Transmembrane mucin

Mucin generic structure

N-terminus

O-glycans

TR

TR

TR

Potential sites for N-glycans

Cytoplasmic tail (binding sites for multiple signalling molecules)

C-terminus

Secretory mucin

D1

D2

D'

D3

CysD

TR

CysD

D4

B

C

CK

vWF-like domain

Leucine zipper
researchers have access to large, relatively organized cancer cases including RNA-sequences (RNA-seq), mutation and aberrant chromatin remodeling, and matched clinical data.

Aside from mucins, several other mutated genes have been ascribed to PC. The most consequential and frequently detected mutation in PC is a constitutively active KRAS mutation contributing to unchecked proliferative signaling via the MAPK pathway or cell growth and anti-apoptotic signaling via the PI3K cascade and observed in over 90% of PC tumors. Abrogation of KRAS signal has been postulated to yield significant benefits to PC patients. Mutations in TP53, observed in approximately 70% of PC tumors, further dysregulate PC cell proliferation. Similarly, the tumor suppressor cyclin-dependent kinase inhibitor 2A (CDKN2A), which codes for p16 and p14arf, is commonly mutated, also contributing to unrestrained cell proliferation.

Relative to survival benefits achieved in other cancers, pancreatic ductal adenocarcinoma (PDAC) continues to rise in both incidence and mortality with a 5-year survival rate of 10%. The underlying causes are multi-factorial, including late diagnosis due to lack of sensitive biomarkers, inadequate therapeutic options, asymptomatic progression, and early metastasis. One of the earliest and most explored events in cancer is rewiring of the cellular pathways subverting physiological gene inactivation. This process is evident in the aberrant expression of mucins in PDAC. The large glycoprotein family of mucins has been widely implicated in many aspects of PDAC biology, including progression, metastasis, and drug resistance. Still, no studies to
date have comprehensively investigated the full expression profiles of mucins in PDAC or their involvement in patient survival.

Further rewiring of cellular pathways is facilitated by the high processivity of genetic information during cancer progression. This leads to the appearance of abundant and novel alternatively spliced mRNA species. Alternative splicing includes many possible modifications to the mRNA transcripts, including exon exclusion or intronic inclusion, due to the use of cryptic splice sites or splice site masking. These modifications alter the functional domains in the translated proteins and contribute to neo-functionality or elimination of regulatory domains. Therefore, this increased proteomic repertoire is hypothesized to impart a substantial benefit to malignant cells. Many of these alternatively spliced variants (SVs) are cancer-specific and may represent reliable biomarkers for prognostic or diagnostic purposes. Investigations of mucin alternative splicing in PDAC and its role in pathology and targeting have been limited to MUC1 and MUC4 SVs with severe truncation of tandem repeats or MUC1 cytoplasmic domain activity. Given the expression of MUC1 in healthy pancreatic tissues, studies of other mucins, including MUC4, MUC5AC, and MUC16 alternative transcripts, might be more impactful in PDAC, as these mucins are expressed explicitly during oncogenesis. Mainly, MUC4 expression is unique to PDAC, notably absent in normal or nonmalignant inflammatory conditions of the pancreas. No studies have been published to date exploring the link between mucin SVs and survival outcomes in PDAC patients.
Another problem related to PDAC is the *one-size-fits-all* clinical approach. The evolving field of cancer therapy has greatly benefited from the identification of cancer subtypes \(^{173-175}\), which has aided in the therapeutic success in several malignancies, with breast cancer undoubtedly being the most impacted \(^{176, 177}\). Subtyping strategies of PDAC tumors have been reported in three principal studies and reflect transcriptomic analysis of whole tumor \(^{178}\) or virtual microdissection \(^{179}\), or genomic analysis \(^{22}\). These approaches, while meritorious for addressing an underexplored area of PDAC, suffer from drawbacks affecting their widespread utilization in the clinic, namely the need for expensive sample processing and significant computational analysis.

### 1.J The Cancer Genome Atlas

Modern advances in cancer discoveries yield substantial power levied from next-generational sequencing technologies combined with matched clinical history. The increase in computing power over the past decade has tremendously empowered our ability to screen entire transcriptomes of many patients to identify putative biomarkers and prognostic indicators. Second to the Human Genome Project, a joint effort by the US National Cancer Institute and the US National Human Genome Research Institute in 2006 resulted in the Cancer Genome Atlas (TCGA). TCGA represents arguably one of the most important databases for cancer bioinformatics ever compiled. Contained within the publicly available TCGA servers is more than 2.5 million gigabytes of genomic, transcriptomic, proteomic, epigenetic, and clinical data from over 20,000 primary tumors and normal tissue samples in 33 cancer types. While many incredible discoveries have already been
described from its data, TCGA is humanity’s first major step towards a collaborative moonshot cure for cancer.

1.K TCGA PDAC Characterization

The initial assessment of the 150 PDAC samples uploaded to the pancreatic cancer dataset (PAAD) to TCGA was conducted by Raphael et al. of The Cancer Genome Atlas Research Consortium. An in-depth analysis of the genomic and transcriptomic profiles of the included patient samples revealed a wealth of information about the cellular alterations occurring in PDAC tumors and is credited with establishing our clinical knowledge about the driver and passenger mutations associated with the disease (Fig 1.11). Over 90% of tumors sequenced demonstrated missense mutations coding for constitutively activated oncogenic KRAS, with G12D being the most prevalent at nearly 45% of detected KRAS alterations. Only 4 patients demonstrated KRAS amplification. Several tumor samples also demonstrated multiple KRAS mutations throughout the tumor sample while few samples were assessed at the single cell level revealing multiple mutations of KRAS in the same cell. Interestingly, in the ten KRAS wild-type samples, other unexpected driver mutations were discovered including GNAS (n=3), CTNNB1 (n=2), and BRAF (n=2). Six of these 10 samples were found to have MAPK pathway activating mutations downstream of KRAS, potentially abrogating the need for KRAS activating mutations. The frequently cancer associated Tp53 was mutated in over 70% of the samples. One of the most influential aspects of this study was the assessment of PDAC tumor cellularity. The tumor ratios were scored for all samples by pathologists upon
Figure 1.11: Genomic Profiling of PDAC Tumors from TCGA

Integrated genomic data for 149 non-hypermutated samples (columns), including mutations (Truncated, In-Frame, or Missense), amplifications and homozygous deletions (“Deep Deletion”), fusions derived from mRNA data, and germline mutations for selected genes as described in the text. Overall number of mutations/Mb and clinicopathologic data for each sample are shown as tracks at the top. Significantly mutated genes (q \% 0.1) from exome sequencing data listed in order of q value, followed by other recurrently altered genes organized in functional classes of oncogenes (Red), DNA damage repair genes (Green), and chromatin modification genes (Blue). Significantly mutated genes from these classes are colored accordingly. The percentage of PDAC samples with an alteration of any type is noted at the left.

FIGURE 1.11: GENOMIC PROFILING OF PDAC TUMORS FROM TCGA

Significantly mutated genes

Significance (−log10(q values))

Genetic Alteration

Amplification Deep Deletion Germline Mutation Fusion

Truncating Mutation Inframe Mutation Missense Mutation

100% 90% 80% 70% 60% 50% 40% 30% 20% 10% 0%
Chapter I: Introduction

submission of sample to the study. The degree of difference between the samples was minimal by pathologist assessment. The team, however, used an *in-silico* method (ABSOLUTE Algorithm) to compute purity scores and revealed a stratification that differed greatly from pathologists (Fig 1.12). The implications of the range of malignant cellularity scores on PDAC next-generation sequencing analyses are far reaching and demand significant consideration in future studies. The TCGA Consortium further assessed tumor purity and gene signatures for PDAC subtyping and compared to three dominant stratification strategies presently used. An exploration of those subtyping methods, their genome or transcriptome signatures, and relationship with tumor sample purity is addressed in the following section.

1.1 PDAC Subtyping

Cancer subtyping has greatly influenced the trajectory of the clinical approach to oncology. Diagnostic, prognostics, and therapeutic interventions are all benefiting from a targeted application to unique subtypes of cancers. As addressed numerous times, the advancement of technologies and computational processing power have significantly empowered researchers’ abilities to begin segregating cancers into subtypes, often with drastically different pathobiology and indicated treatments. Undoubtedly, substantial work lies ahead to begin to elucidate the major aspects of this field, but some cancers are already benefiting immensely from this work. The best example of the clinical progress facilitated by gene expression-based subtyping is breast cancer (BC). Previously, clinical BC typing was based on classification built on the collective physical parameters of
Figure 1.12: Assessment of PDAC Tumor Purity

Samples submitted to TCGA for inclusion with the PDAC dataset were initially screened by pathologists for tumor sample purity. These scores were compared to two different computational methods of assessing tumor cellularity. While Mode methylation was near the quartile range and median as the ABSOLUTE method, its segregation of samples into High and Low purity were similar to that reported by pathologists. In both cases, few samples cross the medians into the opposite purity zone. The ABSOLUTE algorithm was selected as the most reliable predictor of tumor sample purity.

FIGURE 1.12: ASSESSMENT OF PDAC TUMOR PURITY
tumors (dimensions, location, nodal status, grade) and histological assessment (HER2 overexpression, estrogen [ER] and progesterone [PR] receptor status).

Comparing the immunohistochemistry, tumor classifications were established from a ‘diagnostic tree’ yielding four subtypes using luminal+/-(hormone receptor expression) and HER2+/− status, and basal+/− status for luminal-/HER2- tumors (ER-/PR-/HER2- tumors are termed Triple Negative and further subtyped by expression of basal markers). A large cohort meta-analysis uncovered significant differences in survival and responses to therapeutic modalities between the five subtypes. These early investigations into BC subtyping and the elucidation of HER2 overexpression and amplification effects contributed to development of the revolutionary class of HER2/neu antagonist drugs and substantial increase in survival of these affected BC patients. Still, BC subtyping continues to benefit from gene sequencing studies. Sequencing studies comparing gene expression signatures of breast tumors in murine models to human tumors revealed many previously uncharacterized genes with potential targeting applicability and a novel subtype marked determined by low expression of the tight junction family claudin. The claudin-low tumors were found to have epithelial-mesenchymal transition signatures, mammary stem cell-like features, and tumor initiating properties. While investigations into BC expression signatures are showing rapid progression in the characterization of BC tumors and enhancing novel treatment concepts, the use of next-generational sequencing technologies hold great promise across many diseases. Here, I will
review the most well-known subtyping studies into PDAC that extend beyond identification of previously described aggressive genotypes.

1.L.1. Collisson Subtyping

One of the earliest attempts to define selective PDAC patient populations based on transcriptional profiling was conducted by Collisson et al.\textsuperscript{178} This group premises their work by stating that subsets of tumors and the use of specific therapies for each has been successful in treating breast and lung cancers, and the same progress may be achievable in PDAC. Next, they home in on a key issue with research involving PDAC, namely the availability of patient tumor samples is grossly lacking. In an insightful moment, the team acknowledges that they can abrogate the paucity of PDAC samples by using publicly available sequencing data for tumor samples included in other studies to boost their sample. Using a non-negative matrix factorization, they established a 62-gene profile that they then reported identifies three different subtypes (Classical, Quasimesenchymal, and Exocrine-like) based on their interpretation of known biological roles of associated genes (Fig 1.13A). Classical PDAC demonstrated high expression of adhesion-associated and epithelial genes. Quasimesenchymal PDAC showed robust expression of mesenchymal-associated genes. Exocrine-like PDAC had high expression of digestive enzymes. Further, this study demonstrated significant differences in survival of patients classified to each of the three subtypes with Quasimesenchymal PDAC patients having the shortest overall survival and Classical PDAC patients surviving the longest despite resulting in a near linear slope (Fig 1.13B). While a noteworthy and well-received conclusion, this study has
minor oversights that are potentially amplified in the application of its findings. Most importantly, the samples used for microarrays were not assessed for tumor composition likely resulting in the general characterization of the three PDAC subsets. Still, this study serves as an important steppingstone for the study and expansion of PDAC subtyping and successor studies build on the conclusions made in this report.

1.L.2. Moffitt Subtyping

The second major attempt to subtype PDAC, by Moffitt et al., made significant strides to further characterize PDAC gene expression. The central supposition of their attempt laid in the histological observation that the bulk of the PDAC tumors are comprised of stromal components with few tumor cells. Through a series of very robust computational methods using microarray data publicly available for PDAC primary and metastatic tumors, as well as normal pancreatic tissues, the team was able to virtually microdissect gene expression profiles from tumor, stromal, and normal compartments of sequenced PDAC samples. These results, in addition to being a profound achievement, led the researchers to discover gene expression signatures that greatly expanded the subtypes defined by Collisson et al. They were able to further separate Collisson’s Classical PDAC into a second group they called ‘Basal-like’ (Fig 1.14A), the latter of which demonstrate high expression of basal factors similar to basal positive breast cancer. Patients with basal-like PDAC tumor also demonstrated significantly shorter survival, with median life of 11 months following diagnosis compared to 19
Figure 1.13: Colisson Subtyping of PDAC Tumors

Subtypes of PDAC tumors and cell lines and their prognostic significance. (A) Heat map of three subtypes of PDAC in DWD-merged UCSF and GSE15471 PDAC microarray data sets using the PDAssigner gene set. (B) Kaplan-Meier survival curve comparing survival of individuals with classical (red), QM-PDAC (blue) and exocrine-like (green) subtypes. P value is by log-rank test.

FIGURE 1.13: COLLISON SUBTYPING OF PDAC TUMORS
months seen in Classical PDAC patients (Fig 1.14B). This study also subdivided the Collisson Quasimesenchymal into a Normal Stromal Subtype and an Activated Stromal Subtype. The impactful concept behind this approach is that their virtual microdissection was able to extrapolate the gene expression profile of the stroma from the whole tumor sample sequence. This is important because a wealth of studies have shown that the pathology of in situ tumors involve biological activities that extend beyond cancer cells involving the surrounding populations of stromal, immune, and non-cancer cells. Moffitt and group described their activated stromal subtype having high expression of macrophage-associated genes, as well as genes attributed to tumor promotion, such as Wnt, matrixins, and fibroblast activation protein (Fig 1.14C). Patients with Activated Stroma PDAC had a reduced median survival of 15 months, compared to 24 months in Normal Stroma PDAC patients (Fig 1.14D). When the various combinations of stromal and tumor type were merged, an interesting trend in survival was observed (Fig 1.15). Moffitt and team showed that Classical tumors with Normal Stroma have better survival than Classical tumors with Activated Stroma; in fact, the former demonstrated the best survival of all four subtype combinations. However, Basal-like tumors abrogate any survival benefits ascribed to Normal Stroma. This study substantially enhanced our ability to demarcate PDAC biology and begin to conceptualize a personalized approach to therapeutic intervention for patients. Despite the advances made by this report, few minor shortcomings stand out. Principally, the whole foundation of the work relies on virtual dissection based on an algorithm to distinguish normal from disease cells trained using microarray data and
Figure 1.14: Moffitt PDAC Subtyping Establishes Two Tumor and Two Stromal Subtypes

Tumor-specific gene expression suggests two subtypes of PDAC tumors. (A) Consensus-clustered heat map of primary tumors, metastatic tumors, and cell line models of PDAC generated using correlation, with the underlying distance function showing two subtypes of PDAC. (B) Kaplan-Meier survival analysis of patients with resected primary tumors from each tumor subtype in A showing differential prognosis among the subtypes with a hazard ratio of 1.89 (95% CI = 1.19–3.02, \(P = 0.007\)). The dual action of stroma is described by distinct gene expression patterns, which are not present in PDAC cell lines. (C) Consensus-clustered heat map of University of North Carolina (UNC) primary tumor samples, metastases and cell lines generated using genes from stromal factors. Samples clustered into three groups, describing samples with activated stroma, samples with normal stroma and samples with low or absent stromal gene expression. (D) Kaplan-Meier survival analysis of patients with resected PDAC from the activated and normal stromal clusters shows that samples in the activated stroma group have worse prognosis, with a hazard ratio of 1.94 (CI = 1.11–3.37, \(P = 0.019\)).

FIGURE 1.14: MOFFITT PDAC SUBTYPING ESTABLISHES TWO TUMOR AND TWO STROMAL TYPES
Figure 1.15: Combining Moffitt Tumor and Stromal Types Demonstrates Survival Differences

Multivariate survival analysis of tumor and stromal subtypes. (A) Heat map of tumor samples using 25 genes from each of the tumor and stromal factors, with samples sorted horizontally by classification. Signature scores for selected gene sets appear above for each sample. (B) Combined Kaplan-Meier survival analysis of resected primary tumors from patients with basal-like or classical tumors and normal or activated stroma showing differential survival ($P < 0.001$, log-rank test). Differential prognosis among the subtypes shows complementarity. Classical tumors with normal stroma ($n = 24$) had the lowest hazard ratio of 0.39 (95% CI = 0.21–0.73), whereas basal-like tumors with activated stroma ($n = 26$) had the highest hazard ratio of 2.28 (95% CI = 1.34–3.87).

FIGURE 1.15: COMBINING MOFFITT TUMOR AND STROMAL TYPES DEMONSTRATES SURVIVAL DIFFERENCES
subsequently applied to RNA-Sequence datasets. While not intrinsically erroneous, this approach does potentially open the door to improperly derived conclusions, such as the massive difference in sensitivity of the two techniques. Further, given that non-cancer, tumor-associated tissue surrounding the tumor cells are attuned to support tumor development, one might question where the ‘zone of most influence’ begins and ends. In other words, how far from the tumor core does the adapted gene expression profile extend? Despite remaining questions, this study was the first to make major advances in PDAC subtyping and created the measure by which future subtyping strategies might be compared.

1.L.3. Bailey Subtyping

While the previous subtyping studies aspired to describe the characteristics of established PDAC tumors, the next major study sought to define disease initiation routes by gene expression profile. Bailey et al. began with a huge cohort of PDAC tumor samples and derived four entirely novel expression-based subtypes. Their approach employed an unsupervised clustering of 96 samples with greater than 40% tumor content, the results of which were applied to 232 samples. The expression fingerprints were assessed for biological networks to ascertain their features (Fig 1.16A). The first type was highlighted by involvement of inflammation, metabolic remodeling, autophagy regulation, EGF signaling, and the transcription factor Tp63ΔN and its target genes. Bailey and group ascribed the last group of genes to features of breast, bladder, lung, and head and neck cancers, since this transcription factor regulates epithelial genes in the absence of functional Tp53, which led them to term this subtype Squamous PDAC. Next, they
described a subset that indicated activation of genes attributed to embryological development of the pancreas, including PDX1, HNF1A/B, HNF4A/G, FOXA2/3, and HES1. Consequently, they labeled this subtype Pancreatic Progenitor PDAC. Important to this dissertation, they also described Progenitor tumors as having high MUC1 and MUC5AC expression, which they also recognized as IPMN-associated mucins, and which led them to hypothesize that this subtype might arise from IPMN progression. The third group of tumors was enriched for genes activated during functional pancreatic differentiation and defined the endocrine and exocrine roles of pancreatic cell populations, such as digestive enzymes and genes linked to Maturity-Onset Diabetes of the Young (MODY). Thus, this subtype was labeled Aberrantly Differentiated Endocrine Exocrine (ADEX). The final gene signature involved immune cell function pathways, including antigen presentation and TCR signaling pathways. The group called this final subtype Immunogenic PDAC. Assessing the impact of these four subtypes on patient survival, Squamous PDAC demonstrated the worst survival with a median of 13 months with little difference (30 vs 24 vs 26 months) between the other three subtypes (Fig 1.16B). This study is commendable in that it attempts to define biological origins and functional evolution of PDAC tumors and potentially could indicate unique therapies for successful treatment of patients with each type, aside from the significant impact this study has in PDAC subtyping, it does rely on microarray data which, as noted for the previous study, has lower sensitivity than next-generational sequencing. Given that the researchers highlight key transcriptional factor-activated pathways, the reduction in sensitivity has the possibility of obscuring lower expressed factors
Figure 1.16: Bailey PDAC Profiling Establishes Four Subtypes

(A) Molecular classes and transcriptional networks defining PDAC. Unsupervised analysis of RNA-Seq data identified 4 PDAC classes: squamous (blue); ADEX (abnormally differentiated endocrine exocrine; brown); pancreatic progenitor (yellow); and immunogenic (red). *P < 0.05, Fisher’s exact test. (B) Kaplan–Meier analysis of patient survival stratified by class.

FIGURE 1.16: BAILEY PDAC PROFILING ESTABLISHES FOUR SUBTYPES

**A**

- **KRAS**
- **CDKN2A**
- **SMAD4**
- **TP53**
- **KDM6A**
- **MLL3**
- **MLL2**
- **TGFB2**

SV subtypes:
- **BRCA1**
- **BRCA2**

* Histopathology

- Moffitt tumor class
- Moffitt stromal class
- Collisson class

- Squamous
- ADEX
- Pancreatic progenitor
- Immunogenic

- Scattered
- Focal
- Stable
- Unstable

- Basal-like
- Classical
- Activated stroma
- Normal stroma

- Quasi-mesenchymal
- Exocrine-like
- Classical

- Non-silent SNV or indel
- SV
- Deletion
- Amplification (copy number ≥8)

- Adenosquamous carcinoma
- Acinar cell carcinoma
- IPMN with invasion/mucinous
- BRCA2 germline mutations

**B**

- \( P = 0.0302 \)
- Median survival:
  - 30.0 vs 23.7 vs 25.6 vs 13.3 months
- \( n = 93 \)
or expression regulators associated with these transcription factors. Additionally, the presence of endocrine and exocrine expressed genes in the samples is a hotly contested finding and islet and/or acinar cell contamination of their samples has been suggested, as Bailey and group used whole tumor for their study. However, this very recent report had an important role in encouraging the direction of the original work described in this thesis.

1.1.4. Law Subtyping

While the previous subtyping studies were based on genomic and transcriptomic profiling, a recent report from Law et al. employed proteomic technologies to profile the translational repertoire of PDAC tumors.\textsuperscript{186} While RNA-based profiling has the great potential to implicate biological aberrations within PDAC tumor cells, proteomic profiling has the added advantage of defining the active functional signature of disease. Law and team were able to correlate their findings to the previous three PDAC subtyping reports significantly (Fig 1.17A-F). By assessing their identified peptides through Gene Ontology enrichment, they report that they can distinguish Moffitt’s classical, low stromal tumors into two distinct new subtypes marked by expression of either Metabolic reprogrammed or progenitor-like subtypes, the latter term was elected due to the similarities with the Bailey Progenitor-like signature. Law and group continue to describe additional new subtypes which overlapped with Bailey’s Squamous and Collisson’s Quasimesenchymal subtypes. However, based on their proteomic discovery, they further discriminate the Bailey Squamous into two unique subtypes they call
Figure 1.17: Law et al. Subtyping Derives Four New PDAC Subtypes with Overlap of Previous Subtyping Studies

(A) Heatmap of samples showing the association between protein expression and the proteomic subtypes, with proteomic subtypes compared to subtypes from Moffitt, Collisson, and Bailey strategies. The missing data in the ribbon above the heatmap indicate the signatures’ scores for these samples did not reach the threshold to accurately assign a corresponding transcriptomic subtype. The red and blue colors in each pixel indicate protein up- and downregulation, respectively.

(B) Representative signature gene expressions in the Moffitt et al. classification scheme across the four proteomic subtypes. The signature scores of the four proteomics subtypes in the (C) Moffitt et al., (D) Collisson et al., and (E) Bailey et al. classification systems. Kaplan–Meier survival curves of (F) all patients, (H) combined proliferative and inflammatory subtypes, and (I) combined metabolic and progenitor subtypes.

FIGURE 1.17: LAW ET AL. SUBTYPING DERIVES FOUR NEW PDAC SUBTYPES WITH OVERLAP OF PREVIOUS SUBTYPING STUDIES

A

B

C

D

E

F

G

H

I

93
Proliferative and Inflammatory PDAC due to pathway scoring. This report is notable in that the researchers are able to compare their PDAC subtypes to therapeutic outcomes (Fig 1.17E-I). While no significant differences were observed in Proliferative and Inflammatory PDAC patients treated with either gemcitabine or the combination line FOLFIRINOX (leucovorin, fluorouracil, irinotecan, and oxaliplatin), patients with Metabolic and Progenitor-like PDAC demonstrated better response to FOLFIRINOX therapy compared to monotherapy gemcitabine. This study is inimitable as it steps away from transcriptional analyses (commonly criticized for failure or difficulty to attribute to the macro-picture of biological activity in cancers) and goes directly to protein expression. This benefit comes with the tradeoff of sensitivity. While mass spectroscopy can identify thousands of peptide sequences with robust expression, RNA sequencing and microarray platforms can identify tens of thousands of RNA species even at very low concentrations. Additionally, this study used PDAC tumor samples isolated from PDAC liver metastases. Therefore, the conclusions of this work must be limited to the academic scope of disease signatures involving suitability of tumor and metastatic site adaptations. Even more, the samples used in Law’s work were whole tumor specimen and hepatic cell contamination cannot be entirely ruled out.

1.M Conceptualization of Dissertation Work

Based on this extensive literature, this dissertation work sought to achieve five major accomplishments. First, a complete mucin expression profile for PDAC has never been published despite the indisputable involvement of mucins in PDAC pathology, initiation, treatment response, and progression. Therefore, it is fitting to
begin with a complete assessment of mucin transcription using publicly available PDAC RNA-Sequencing datasets. Next, this dissertation research aspired to marry the concepts of cancer-associated alternative splicing with mucin expression, describing the expression and prognostic implications of alternatively spliced mucin species. Thirdly, because mucins have biological involvement, this work sought to derive gene expression signatures from unique mucin clusters to explore the possibility of a novel mucin-based PDAC subtyping strategy. Fourth, in this research, I selected a significant prognosis-associated mucin splice variant, and determined its biological function contributing to more aggressive PDAC disease. And finally, using the sequence of prognostic mucin splice variants, I aimed to develop a novel diagnostic assay using patient bio-fluids, and to explore its potential clinical utility.
**Hypotheses and Dissertation Aims**

The overall goal of this dissertation is to explore the expression of Mucins (MUCs) and their alternative splice variants (SVs) in PDAC. Pancreatic cancer (PC) has a 5-year survival rate of only 10% and is projected to increase in both incidence and mortality. This grim perspective of PC is likely the consequence of a multitude of converging factors, including a lack of early symptoms contributing to late diagnosis, early metastasis, and poor understanding of its pathology. With no reliable diagnostic or prognostic biomarkers and the lack of actionable targets, the next significant clinical milestone toward better treatment of PC patients is the identification of PC-specific biomolecules that can assist in earlier diagnosis and therapeutic targeting. MUCs have gained significance for their role in PC progression, diagnostics, and therapeutic targeting. However, their precise mechanistic contribution is poorly understood.

Based on preliminary studies and published literature, the central hypotheses of this thesis are 1) MUCs are selectively expressed in PDAC patients, 2) MUC SVs alter their typical functions in expressing cells and contribute to pancreatic cancer pathology, and 3) MUC SVs may serve as unique diagnostic and prognostic markers detectable from patient plasma. I proposed the following specific aims to address these hypotheses.

**Specific Aim I:** Surprisingly, a composite dissection of MUCs expressed in human PDAC patient tumors has not been reported. Further, alternative splicing of PDAC-associated MUCs has yet to be evaluated. Due to the abundance of typical and aberrantly expressed MUCs in PDAC, I hypothesized that mucin profiles might
highlight a new avenue of PDAC research. Utilizing RNA-sequencing data from 150 PC patients in the TCGA database, total expression of mucins and their annotated SVs were explored and compared to disease parameters. Additionally, all known MUCs and MUC SVs were correlated with the survival outcome in TCGA PDAC patients. Using the TCGA PDAC dataset, I established four clusters of MUCs and correlated them with activated pathways indicated by specific gene expression correlated to MUC principal component analysis, identifying four putative novel PDAC subtypes.

**Specific Aim II: Explore the functional consequences of MUC4Δ6 in PC disease pathogenesis.**

Examination of the MUC4Δ6 sequence revealed that its processing produces an in-frame deletion of exon 6 interfering with its nidogen-1-like (NIDO) domain. Further investigation using PCR demonstrated that MUC4Δ6 is positively and exclusively expressed in PC cell lines and patient surgical resection samples. Previous studies from our group have investigated the pathological implications of MUC4 SVs lacking tandem repeat domains in MUC4/X and MUC4/Y and the importance of MUC4 NIDO domain in metastasis of PC. However, the functional relevance of MUC4Δ6 in the presence of wild-type MUC4 in PDAC has not been studied. I hypothesized that MUC4Δ6 cooperates with wild-type MUC4 to promote invasion and migration of PDAC cells. Given that expression of MUC4Δ6 corresponds to poor survival time in PC patients and loss of exon 6, which corresponds to the NIDO domain and confers stromal interaction functionality,
plausibly it induces altered or neofunctionality in the presence of wild-type MUC4. Therefore, I generated an overexpression vector and transfected it into PDAC cell lines having intrinsic MUC4 expression to explore the functional effects of MUC4Δ6 expression. I planned to compare the in vitro differences in cell proliferation, viability, and invasion and in vivo tumorigenesis and metastasis in a mouse model.

**Specific Aim III: Validate the detection of mucin SV in patient plasma while confirming their specificity for PC.**

I hypothesized that PC-specific, alternatively spliced and mutated gene transcripts can be detected in patient plasma and that my predicted MUC4Δ6 SV from Aim I is a unique marker of PDAC. Therefore, I developed gold nanoparticles covalently linked to short sequence-specific oligoprobes with conditional fluorescence requiring target hybridization and enzymatic cleavage for signal activation. My nanoparticle assay overcomes the hurdles of specificity of antibodies, spatiotemporal expression, stability of protein of interest, and ease of multiplexing. Since the action of DNA duplex-specific nuclease requires precise basepair complementarity, my nanoprobe assay can serve as a plasma-based method to detect numerous spliced and mutated transcripts. Based on the prognostically significant MUC4Δ6 from my analysis of TCGA patient data, I demonstrated using in vitro methods that this transcript is: 1) detectable in patient plasma and 2) specific to PC in a disease stage-independent manner. I tested my novel nano-probe assay using samples from our lab’s extensive library of patient plasma samples derived from malignant, benign, and nonpathological conditions, detecting the MUC4Δ6 RNA in circulating RNA using my novel assay. In addition
to MUC4Δ6 probes, I demonstrated using specific probes that I can also detect PC-specific and constitutively active KRAS$^{G12D}$.
CHAPTER 2:

MUCIN EXPRESSION PROFILE IN PANCREATIC DUCTAL ADENOCARCINOMA
2.A Introduction

Here, I present a focused bioinformatics-based strategy using PDAC RNA-Seq data from TCGA (gdc.cancer.gov) for the expression of mucin genes and SVs. I investigated the importance of assessing the cellularity of PDAC tumor samples when analyzing transcriptomic data and demonstrated a quantitative strategy to correct mucin expression based on malignant cellularity. I utilized the extensive expression of mucins to explore correlations between specific mucin genes. These correlations were used to identify four mucin-based subtypes and explore the possible biological processes activated within each specific PDAC subtype. Finally, I assessed the expression of all 107 known mucin SVs and the impact of their expression on the overall survival outcome. I validated the expression and survival association of two opposing SVs. Taken together, our analyses suggest that selective expression of mucins may have novel implications in PDAC tumor biology, and certain SVs may contribute to disease parameters, impact the survival of PDAC patients, and serve as excellent diagnostic and prognostic markers. Clinical advances in next-generation profiling to identify novel therapeutic targets or diagnostic biomarkers in PC lags behind other malignancies while treatment outcomes remain unimproved. This disparity in therapeutic response is partly due to numerous physical characteristics of PC tumors, including exacerbated desmoplastic reaction, poor drug perfusion, innate therapeutic resistance, and further complicated by the majority of diagnoses being made in late-stage disease. We and others have demonstrated that PC tumors aberrantly express abundant mucins and...
generate a large number of alternatively spliced isoforms \cite{129, 137-141, 188, 189}. Experiments on many mucin-associated domains have provided limited evidence for functional interactions. Though molecular studies have revealed several consequential alterations in PC, like the expression of constitutively active KRAS mutant, loss of cell cycle checkpoint inhibitors, and aberrant mucin expression, the complete story remains obscured. I believe that the results of this project address the shortcomings in available scientific knowledge, including: 1) if PC tumors express isoforms specific to the disease setting, 2) detectability of disease-associated splice variants from biofluids, 3) the prognostic implication of alternatively spliced genes in the course of disease progression, and 4) functional consequences of alternatively spliced genes.

2.B Results

2.B.1. General outline of patient samples

The pancreatic cancer dataset (PAAD) in TCGA contains a total of 182 patient samples. Of those, 8 were later confirmed neuroendocrine, 24 were other diagnoses (acinar carcinoma, unknown origin, etc.), and 150 were histologically confirmed PDAC (Fig 2.1A). Of those 150 PDAC, all had cellularity scores and applicable clinical history except one. Almost 83% of those 149 samples were diagnoses between the ages of 50-80 years (Fig 2.1B) and over half were male (Fig 2.1C). Comparing the histological stage, approximately 50% were Grade 2, 46% were Grade 3, while 5 were Grade 1 and a single samples was Grade 4 (Fig 2.1D).
Figure 2.1: TCGA PDAC Patient Demographics

Patients in the Pancreatic Cancer Dataset (PAAD) of TCGA are associated with a variety of demographic and clinical parameters. (A) The majority of patients were between the ages of 50-80 years old at the time of diagnosis. (B) Of the 182 total cases, only 150 were confirmed PDAC after submission and upload. This population was selected for analyses. (C) Just over half the PDAC cases were male despite no apparent increased sex-based risk of developing PDAC in literature. (D) Out of the 150 total PDAC cases, 149 cases had tumor sample purity scores reported. Around half of these cases were Grades 2 and 3.
FIGURE 2.1: TCGA PDAC PATIENT DEMOGRAPHICS

A

Number of Subjects

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-40</td>
<td>1</td>
<td>14</td>
<td>29</td>
<td>48</td>
<td>46</td>
</tr>
<tr>
<td>40-50</td>
<td>8</td>
<td>24</td>
<td>75</td>
<td>149</td>
<td></td>
</tr>
</tbody>
</table>

B

Number of Subjects

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroendocrine</td>
<td>8</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
</tr>
<tr>
<td>PDAC</td>
<td>150</td>
</tr>
</tbody>
</table>

C

N=69

N=80

D

PC

Grade 3: 68

Grade 2: 75

Grade 1: 12

Grade 0: 1

Prior Treatment: Undifferentiated

1% Cellularity: 1

Acinar Cell Carcinoma: 8

Adenocarcinoma Normal: 9

Adenocarcinoma Aden Normal: 3

Non-pancreatic Origin: 3

MPMN: 8

Metastatic: 3

Neuroendocrine
2.B.2. Mucins expression influenced by malignant cellularity

Characterization of patients’ samples in TCGA dataset revealed robust expression of the epithelial marker CK19 in high cellularity samples (determined by individual ABSOLUTE Purity scores above the median of all samples reported by the TCGA Consortium on the PAAD dataset). However, samples demonstrated a wide variation in amylase expression, with high cellularity expressing 0-23787 transcript per million (TPM, median= 3.3 TPM) and 0 to 41320 TPM (median= 41.4 TPM) in low cellularity. (Fig 2.2). Low expression of CD45 and leptin in both high and low cellularity groups indicated low immune and adipocyte populations. Mesothelin expression was unsurprisingly substantial, with a median expression of 326.9 TPM and 132.8 TPM in high and low cellularity, respectively.

Mucin expression was compared between high cellularity (n=74) and low cellularity samples (n=75) to healthy tissues or between high and low cellularity samples (Fig 2.3). All cases demonstrated a wide range of mucin expression between patients with substantial differences in high and low cellularity groups. Four groups of mucins were established using Pearson’s correlation and five clusters of patients generated by Spearman correlation dendrograms of log-transformed mucin expression values. MUC19 clustered alone (group 1), MUC1/3/12/13/17/20 (group 2), MUC6/15/22 (group 3), and MUC2/4/5AC/5B/16/21 (group 4). Survival analysis of the high cellularity patients based on patient clustering of this mucin expression demonstrated significant differences in overall survival (Fig 2.3B, Wilcoxon p=0.05). Cluster 5 patients demonstrated the highest expression of group 2 mucins and had the longest
Figure 2.2: Cellular Composition of TCGA PDAC Samples

Samples from the dataset indicate high expression of ductal marker keratin-19 (CK19) and stromal component mesothelin (MSLN). Significant differences are observed between high and low tumor purity samples for all genes except acinar-associated amylase (AMY2A). Boxplots were given by log-transformed TPM expression. Significance results are from the student’s t-test.
FIGURE 2.2: CELLULAR COMPOSITION OF TCGA PDAC SAMPLES

Low Cellularity (n=72)  High Cellularity (n=76)
Figure 2.3: Expression of Mucins in the TCGA PDAC Dataset

Expression of mucins in the TCGA PDAC dataset before cellularity correction of the samples shows significant survival differences between the patient of different mucin signatures. (A) Heatmap and boxplots of cellularity-uncorrected expression of all mucins in the combined patient population or stratifying patient samples by malignant cellularity into high (n=74) or low (n=75) cellularity. Four groups of mucin expression profiles were established by correlation dendrogram comprised of MUC19 alone (group 1), MUC1/3/12/13/17/20 (group 2), MUC6/15/22 (group 3), and MUC2/4/5AC/5B/16/21 (group 4), and patients were stratified into five clusters. (B) Kaplan-Meier survival plots of the high cellularity subjects in the five clusters demonstrated a statistically significant difference in early deaths (Wilcoxon p-value=0.05) between the groups with cluster 3 (high group 3 mucins, especially MUC15, and high MUC5B expression) having the worst median overall survival (308 days); cluster 4 demonstrating a median survival time of 394 days; cluster 1 median survival was 460 days; cluster 2 median survival was 545 days; and cluster 5 (highest expression of group 2 mucins) having the longest median survival time (732 days). Significance results are from Wilcoxon test between High and Low Cellularity groups. Mucins were clustered in 4 groups by Pearson correlation dendrogram into 5 clusters of patients by Spearman correlation. All boxplots were given by log-transformed TPM expression. Significance indicated by: *= ≤0.05 and **= ≤0.005.
median survival of 732 days compared to the worst surviving patients (median of 308 days) in cluster 3, which had high group 3 mucins (especially MUC15) and high MUC5B expression.

Further, I assessed the clustering of mucin expression and their association with overall survival in PDAC patients using a secondary RNA-Seq dataset (PACA-AU) from the International Cancer Genome Consortium (ICGC), and confirmed harmonization in the mucin groups identified between both datasets. In this dataset, the four mucin clusters (Fig 2.4A) were comprised of MUC12/22 (group 1), MUC4/15/16/21 (group 2), MUC3A/6/19 (group 3), and MUC1/2/5AC/5B/13/17/20 (group 4). Comparing this dataset to cellularity uncorrected TCGA samples, two groups of mucins showed substantial synchrony. Group 2 mucins of ICGC overlapped TCGA group 4 by MUC4/16/21, and ICGC group 4 was similar to TCGA group 2 with MUC1/13/17/20. Interestingly, although the patient cohort in this dataset clinically resembled TCGA patients, mucin expression was lower than that observed from TCGA. The ICGC patients were clustered by Spearman’s correlations into 5 distinct clusters. Accessing overall survival outcomes in these clusters resulted in significantly different (p=0.02) loss of life (Fig 2.4B). Clusters 4 and 5 had the worst survival with medians of 388 and 361 days, respectively. Cluster 4 patients demonstrated high expression of group 2 mucins while cluster 5 patients had higher group 3 mucins. Both had low group 4 mucin expression. Clusters 1 and 2 both demonstrated similar mucin signatures, including high expression of group 4 mucins, with the exception that cluster 1 also displayed high group 2 expression. The differences in survival were substantial,
Expression and clustering of mucins in the ICGC PACA-AU PDAC dataset demonstrated significant survival differences between mucin signature clusters. (A) Four groups of mucins were identified by Pearson correlation using scaled expression and five clusters of patients sorted using Spearman’s correlation of mucin expression. Mucin groups were comprised of MUC12/22 (group 1), MUC4/15/16/21 (group 2), MUC3A/6/19 (group 3), and MUC1/2/5AC/5B/13/17/20 (group 4). (B) Overall survival in the five patient clusters resulted in significantly different outcomes (Log Rank p=0.02). Clusters 4 and 5 had the worst survival with medians of 388 and 361 days, respectively. Clusters 4 patients demonstrated high expression of group 2 mucins while cluster 5 patients had higher group 3 mucins. Both had low group 4 mucin expression. Clusters 1 and 2 both demonstrated similar mucin signatures, including high expression of group 4 mucins, with the exception that group 1 also displayed high group 2 expression. The differences in survival were substantial, with cluster 1 patients having a median overall survival of 709 days compared to 1144 days in cluster 2 patients.
FIGURE 2.4: EXPRESSION OF MUCINS IN THE ICGC PDAC COMPARISON DATASET AND SURVIVAL ASSOCIATIONS

A

ICGC PACA-AU
MUC Expression

Group 1
-2 -1 0 1 2
MUC12
MUC22

Group 2
MUC21
MUC4
MUC16
MUC15

Group 3
MUC3A
MUC6
MUC19
MUC17
MUC2

Group 4
MUC5AC
MUC5B
MUC1
MUC13
MUC20

Expression (log)
(n=67)

Subject Cluster 1
Subject Cluster 2
Subject Cluster 3
Subject Cluster 4
Subject Cluster 5

B

% Surviving

Cluster 1 (n=10)
Cluster 2 (n=20)
Cluster 3 (n=7)
Cluster 4 (n=16)
Cluster 5 (n=14)

LogRank p=0.02

OS (Days)
with cluster 1 patients having a median overall survival of 709 days compared to 1144 days for cluster 2 patients. The expression of MUC4/16 is likely an important factor underlying this discrepancy. Cluster 3 patients displayed lower mucin expression and had a median survival of 768 days.

The malignant cellularity-corrected expression of mucins significantly adjusted all observed medians (Table1), and four groups of mucin genes in five clusters of patients were identified by standardized correlation heatmap visualization (Fig 2.5A). The mucins sorted into MUC7/12/17 (group 1), MUC1/3/13/19/20 (group 2), MUC6/15/22 (group 3), and MUC2/4/5AC/5B/16/21 (group 4). The overall survival of these five cellularity-corrected clusters of patients (Fig 2.5B) demonstrated a significant difference in early deaths (Wilcoxon p=0.03). Cluster 1 patients appeared to have lower expression of group 2 mucins and strong expression of group 4 mucins (especially MUC16), and the shortest median survival of 293 days. Cluster 2 patients, however, had the longest overall survival of 1059 days, demonstrating higher group 2 mucin expression (particularly MUC1). Clusters 3 and 4 had similar median survival times (598 and 593 days, respectively), with the former demonstrating low group 2 mucin expression but high MUC4 expression and the later exhibiting higher group 2 and group 1 mucin expression. Cluster 5 patient median survival was 738 days and demonstrated a high MUC5AC expression but also higher group 2 mucins (especially MUC13).

Assessing individual mucins, MUC1 expression was highest before and after the cellularity correction. The expression of MUC2/7/19/21/22 was minimal with uncorrected and corrected medians below 1 TPM. The MUC13 demonstrated the
Median expression of all mucins between adjacent normal and PC samples was calculated in the combined patient population and stratified by cellularity in the TCGA dataset and compared to median expression observed in the ICGC dataset. Values reported after TPM normalization.

### TABLE 1: MEDIAN MUCIN EXPRESSION BETWEEN STRATIFIED TCGA PDAC AND ICGC SAMPLES.

<table>
<thead>
<tr>
<th>Sample Stratification</th>
<th>MUCI</th>
<th>MUC2</th>
<th>MUC3</th>
<th>MUC5</th>
<th>MUC6</th>
<th>MUC7</th>
<th>MUC8</th>
<th>MUC9</th>
<th>MUC16</th>
<th>MUC17</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TCGA</strong> Low Cellularity</td>
<td>0.23</td>
<td>0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>TCGA</strong> High Cellularity</td>
<td>0.23</td>
<td>0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>ICGC</strong> Low Cellularity</td>
<td>0.23</td>
<td>0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>ICGC</strong> High Cellularity</td>
<td>0.23</td>
<td>0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Expression and clustering of mucins in the TCGA PDAC dataset after cellularity correction of the samples. (A) Heatmap of cellularity-corrected expression of all mucins and expression boxplots of the combined patient population and after stratifying patient samples by malignant cellularity into high (n=74) or low (n=75) cellularity. Mucin expression correlations resulted in four new profiles comprised of MUC7/12/17 (group 1), MUC1/3/13/19/20 (group 2), MUC6/15/22 (group 3), and MUC2/4/5AC/5B/16/21 (group 4). (B) Kaplan-Meier survival plots of the 5 subject clusters demonstrated a statistically significant difference in early deaths (Wilcoxon p-value=0.0346) between the groups with cluster 1 (lower expression of group 2 mucins and strong group 4 expression, especially MUC16) having the worst median overall survival (293 days); cluster 3 (low group 2 expression and higher MUC4 expression) and cluster 4 (high groups 1 and 2 mucins) having similar median survival times (598 and 593 days, respectively); cluster 5 (high MUC5AC and group 2 expression, especially MUC13) having a median survival time of 738 days; and cluster 2 (high group 2 mucin expression, particular MUC1) having the longest overall survival (1059 days). Significance results are from Wilcoxon test between High and Low Cellularity groups. Mucins were clustered in 4 groups by Pearson correlation dendrogram into 5 clusters of patients by Spearman correlation. All boxplots were given by log-transformed TPM expression. Significance indicated by: *= ≤0.05 and **= ≤0.005.
FIGURE 2.5: CELLULARITY-CORRECT EXPRESSION OF MUCINS IN TCGA GENERATES FOUR NOVEL CLUSTERS
second highest expression level, which significantly increased after cellularity-correction. Among some mucins, cellularity correction reversed the trends in expression. For instance, MUC4 expressed higher in the high cellularity; however, after correction, low cellularity demonstrated the highest expression, suggesting that the malignant cells with higher stromal contents may express more copies of MUC4 compared to the former. However, the higher expression of MUC5AC in high cellularity cases was reaffirmed after cellularity-correction, possibly due to a paracrine- or autocrine-mediated forward signal. I concluded from these analyses that a qualitative stratification strategy would fail to provide an accurate comparison of samples with different cellularity scores within the same strata. Thus, I elected to proceed with the cellularity-corrected mucin expression levels for future analysis.

2.B.3. Mucin-dependent PDAC patient subtyping

Principal component analysis (PCA) of mucin expression using Spearman correlations elucidated mucin signatures of PDAC. The PCA of uncorrected TCGA mucin expression was compared to PCA of ICGC samples. Using the cutoff values explained in the methods section, both datasets resulted in three significant PCs with substantial overlap in dominant mucins contributing to the groups (Fig 2.6A-E). After confirming significant overlap between the two datasets, I proceeded to PCA calculations using malignant cellularity corrected TCGA mucin expression. The PCA resulted in four significant principal components (PC), determined using $p \leq 0.05$ and eigenvalues $\geq 1$, which explained nearly 70% of the variation in expression. The heaviest loading mucins from each of the four PCs were
PCA clustering of mucins overlaps between TCGA and ICGC and suggests mucin signature subtypes are linked to activation of unique pathways. After calculating independent PCAs of mucin expression in the cellularity-uncorrected TCGA and ICGC datasets, the heaviest loading mucins to each principal component were compared. (A) In PC1, all mucins observed in the TCGA dataset were also part of PC1 from ICGC with the addition of MUC6 in the latter. (B) PC2 in both sets contained the heaviest loads from MUC4, MUC16, and MUC20. ICGC PC2 also included a positive contribution from MUC12 and negative contribution from MUC6, while TCGA PC2 variation was additionally explained by positive contribution from MUC5B and negative contributions from MUC3 and MUC17. (C) Variation in PC3 from both datasets was mostly contributed from positive loads from MUC12 and negative contributions from MUC5B and MUC6. TCGA PC3 variation was additionally explained by positive loading from MUC4 and MUC16, while ICGC PC3 variation was further explained by positive contributions from MUC13 and MUC17 and negative load from MUC5AC. (D) Scree plot of PCA on TCGA PDAC cellularity-uncorrected mucin expression shows 3 significant PCs. (E) Scree plot of PCA on ICGC PACA-AU PDAC mucin expression shows 3 significant PCs. (F) From the significant principal components, the top four identified unique MUC expression patterns, summarized within the table, and their load and contribution to the PC.
FIGURE 2.6: PCA CLUSTERS OF MUCINS IN TCGA AND ICGC OVERLAP

A: PC1

B: PC2

C: PC3

D: TCGA Uncorrected PCA

E: ICGC PACA-AU PCA

F: TCGA Corrected PC Loads

<table>
<thead>
<tr>
<th>Mucin</th>
<th>Load</th>
<th>Partial Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC3</td>
<td>0.788</td>
<td>17.039</td>
</tr>
<tr>
<td>MUC1</td>
<td>0.772</td>
<td>16.328</td>
</tr>
<tr>
<td>MUC13</td>
<td>0.728</td>
<td>14.514</td>
</tr>
<tr>
<td>MUC20</td>
<td>0.641</td>
<td>11.283</td>
</tr>
<tr>
<td>MUC12</td>
<td>0.617</td>
<td>10.445</td>
</tr>
<tr>
<td>MUC5B</td>
<td>0.558</td>
<td>8.541</td>
</tr>
<tr>
<td>MUC17</td>
<td>0.517</td>
<td>7.337</td>
</tr>
<tr>
<td>MUC16</td>
<td>0.767</td>
<td>34.782</td>
</tr>
<tr>
<td>MUC17</td>
<td>-0.636</td>
<td>23.964</td>
</tr>
<tr>
<td>MUC4</td>
<td>0.550</td>
<td>17.909</td>
</tr>
<tr>
<td>MUC5B</td>
<td>0.367</td>
<td>7.981</td>
</tr>
<tr>
<td>MUC3</td>
<td>-0.323</td>
<td>6.176</td>
</tr>
<tr>
<td>MUC6</td>
<td>0.629</td>
<td>29.616</td>
</tr>
<tr>
<td>MUC5B</td>
<td>0.499</td>
<td>18.622</td>
</tr>
<tr>
<td>MUC12</td>
<td>-0.460</td>
<td>15.829</td>
</tr>
<tr>
<td>MUC5AC</td>
<td>0.440</td>
<td>14.471</td>
</tr>
<tr>
<td>MUC4</td>
<td>-0.387</td>
<td>11.199</td>
</tr>
<tr>
<td>MUC5AC</td>
<td>0.694</td>
<td>51.236</td>
</tr>
<tr>
<td>MUC6</td>
<td>-0.466</td>
<td>23.183</td>
</tr>
</tbody>
</table>
summarized into distinct panels (Fig 2.6F). I selected mucins with load values $>|0.5|$ in principal component 1 (PC1) and $>|0.3|$ for all others. The PC1 demonstrated the complex association of gene expression comprised of positive contributions from MUC1/3/5B/12/13/17/20. The PC2 was comprised of positive association by MUC4/5B/16 and negatively from MUC3/17. The PC3 encompassed positively from the expression of MUC5AC/5B/6 and negatively from MUC4/12. The PC4 was composed of positive expression of MUC5AC and negatively for MUC6 expression.

I calculated the Spearman correlations of all genes to each of the four PC’s relative to their medians and considered genes that had significant Bonferroni-corrected $p$-values for all correlation groups followed by Ingenuity Pathway Analysis for four positively correlated lists of genes (Fig 2.7). An immune response signature is the most implicated feature in PC1. Based on this finding, I analyzed the expression of CD45 in PC1 patients and found a high expression of CD45 in high PC1 compared to low PC1 (13.9 TPM vs. 7.2 TPM, $p=0.0113$). Pathways analysis of PC2 correlated genes indicated signaling through glucocorticoid receptors, vitamin D, and epithelial-mesenchymal transition (EMT) regulation. Studies have suggested that glucocorticoids can induce EMT, therapy resistance and enhance metastasis in PDAC \textsuperscript{192}; therefore, considering the well-known role of EMT in cancer biology, I designated the PC2 cohort of tumors as progressing and aggressive. The PC3-associated pathways revealed activation of the SPINK1, maturity-onset diabetes of young (MODY), and acute phase response pathways suggesting high acinar involvement and stress-induced responses within the
PCA clustering of mucins overlaps between TCGA and ICGC and suggests mucin signature subtypes are linked to activation of unique pathways. The positively correlated genes from the entire transcriptome to each PC were fed into pathway analysis software to explore the pathways affected. PC significance was determined using Bartlett’s test $\leq 0.05$ and eigenvalues $\geq 1$. Loading thresholds were set to absolute values of 0.5 for PC1 and 0.3 for all other PCs.
FIGURE 2.7: MUCIN SIGNATURES DEMARCATE UNIQUE PATHWAY ACTIVATIONS

MUC-based Subtype Upregulated Pathways

- Granulocyte Adhesion and Diapedesis: 18/180
- Agranulocyte Adhesion and Diapedesis: 18/193
- Melatonin Degradation I: 10/60
- LPS/IL-1 Mediated Inhibition of RXR Function: 19/224
- PXR/RXR Activation: 10/65
- Primary Immunodeficiency Signaling: 8/50
- Comm b/w Innate-Adaptive Immune Cells: 10/96
- Glucocorticoid Receptor Signaling: 11/336
- Protein Citrullination: 2/5
- Reg of EMT Pathway: 6/192
- VDR/RXR Activation: 4/78
- Reg of EMT In Development Pathway: 4/84
- Synaptogenesis Signaling Pathway: 6/312
- Retinoic Biosynthesis I: 1/4
- Reg of EMT By Growth Factors Pathway: 4/188
- Intrinsic Prothrombin Activation Pathway: 2/42
- SPINK1 Pancreatic Cancer Pathway: 15/60
- FXR/RXR Activation: 13/126
- Acute Phase Response Signaling: 11/179
- LXR/RXR Activation: 9/121
- Coagulation System: 5/35
- MODY Signaling: 6/69
- PXR/RXR Activation: 6/65
- Retinol Biosynthesis: 5/42
- Sperm Motility: 12/223
- Nicotine Degradation III: 6/57
- MSP-RON Signaling Pathway: 6/68
- Melatonin Degradation I: 6/60
- Intrinsic Prothrombin Activation Pathway: 5/42
- Nicotine Degradation II: 6/65
- Superpathway of Melatonin Degradation: 6/65
- Serotonin Degradation: 6/67
- Agranulocyte Adhesion and Diapedesis: 9/193
- Reg of Cell Mech. by Calpain Protease: 5/65
- FXR/RXR Activation: 7/126

# Molecules Upregulated

-log(p-value)
tumor. Therefore, I hypothesize that the PC3 associate with a gene signature involving acinar insult or chronic pancreatitis/inflammation. Supporting this assumption, amylase expression in high PC3 patients had a median of 53.5 TPM, and a median of 1.02 TPM (p=0.0306) in low PC3, indicating the significant contribution by acinar cells in the high PC3 patient samples. Many processes correlated with PC4-associated genes imply a downregulation of inflammation, as suggested by the upregulation of melatonin, nicotine, and serotonin degradation. Interestingly, an upregulation of signaling through the c-MET family receptor tyrosine kinase Recepteur d’Origine Nantais (RON) via its ligand Macrophage Stimulating Protein (MSP) suggests a potential route of tumorigenesis through intraepithelial neoplasia (PanINs) progression to PDAC for PC4. Further, naïve Bayes calculations resulted in highly accurate prediction of patients to an Immune Activated/Low Cellularity Tumors (PC1) with an area under the curve (AUC) of 0.9648 (Fig 2.8A) as well as Progressive/Aggressive Transforming Tumors (PC2) with an AUC of 0.9971 (Fig 2.8C). Predictions were slightly less powerful, although still high for Pancreatitis/Acinar Initiated Tumors (PC3) with an AUC of 0.8833 (Fig2.8B) and Anti-Inflammatory/PanIn Initiated Tumors (PC4) with an AUC of 0.8529 (Fig 2.8D). Seeking to explore the impact of mucin-based PDAC subtypes on survival outcome, the TCGA samples were stratified into groups based on their PC score relative to the median, comparing the top 25% (high PC) to the bottom 25% (low PC). Overall median survival outcome was statistically significant for PC1 (p=0.04), with higher PC surviving longer than lower PC (738 vs. 511 days, Fig 2.8E). Survival curves of High and Low PC4 patients indicated a trend toward
Figure 2.8: PCA-Significant Mucins Independently Predict Association to MUC-Based Subtypes

Using the heaviest weighted mucins from the PCA, mucin panels were established and used to predict patient PDAC tumor subtype. (A) Using the cellularity-corrected expression of MUC1/3/5B/12/13/17/20, we were able to accurately predict patients as having immune activated tumors (PC1) with an AUC 0.9648. (B) The cellularity-corrected expression of MUC3/4/5B/16/17 most accurately predicted patient tumors to a progressive/increasing aggressiveness subtype (PC2) with an AUC of 0.9971. (C) The cellularity-corrected expression of MUC4/5AC/5B/6/12 predicted patients to be in the pancreatitis/acinar cell damage induced subtype (PC3) with an AUC of 0.8833. (D) Cellularity-corrected MUC5AC/6 could predict patient tumor subtype anti-inflammatory (PC4) with an AUC of 8529. (E) Comparing survival in the top 25% to the bottom 25% of immune activation subtype (PC1) highlighted a significant difference in overall survival, with high PC1 patients surviving 227 days longer on average (p=0.0497). (F) Comparing the survival of the top 25% to the bottom 25% of anti-inflammatory subtype (PC4), an appreciable trend is observable, with patients having low PC4 scores surviving an average of 172 days longer (p=0.2018).
FIGURE 2.8: PCA-SIGNIFICANT MUCINS INDEPENDENTLY PREDICT ASSOCIATION TO MUC-BASED SUBTYPES

A. Immune Activation

B. Progressive

C. Pancreatitis Initiated

D. Anti-Inflammatory/PanIn Initiated

E. Immune Activation Survival

F. Anti-Inflammatory/PanIn Initiated Survival

High PC (n=37)
Low PC (n=37)

p=0.04
p=0.12
Chapter II: Mucin Expression Profiling

**Figure 2.9: Mucin Splice Variants Correlate with Differential Survival Outcome in PDAC Patients**

Survival plots based on expression above (High Expression) or below (Low/No Expression) the non-zero median of the 9 highest expressed mucin SVs significantly correlated with survival outcome. (A) High expression of four MUC1 and one MUC13 SVs were associated with better survival. (B) Boxplots of these five SVs associated with better survival demonstrate that MUC13-sv-205 has the highest expression, followed by MUC1-sv-222, with the other three MUC1 SVs expressed at similar lower levels. (C) Higher expression of both MUC4, MUC15, MUC16 SVs were associated with worse survival. (D) Boxplots of these four SVs associated with worse survival demonstrate that both MUC4, MUC15, MUC16 SVs expressed at similar levels with MUC16-sv-201 expressed slightly higher. Expression of significant SVs given in log transformed TPM.
FIGURE 2.9: MUCIN SPLICE VARIANTS CORRELATE WITH DIFFERENTIAL SURVIVAL OUTCOME IN PDAC PATIENTS

(A) MUC1-210
High Expression (n=73)
Low/No Expression (n=74)
Wilcoxon p=0.0135*

MUC1-221
High Expression (n=70)
Low/No Expression (n=77)
Wilcoxon p=0.0135*

(B) Expression (Log TPM)

MUC1-222
High Expression (n=72)
Low/No Expression (n=75)
Log-Rank p=0.022*
Wilcoxon p=0.032*

MUC1-226
High Expression (n=74)
Low/No Expression (n=73)
Wilcoxon p=0.038*

(C) MUC13-205
High Expression (n=72)
Low/No Expression (n=75)
Log-Rank p=0.015*
Wilcoxon p=0.015*

(D) Expression (Log TPM)

MUC4-215
High Expression (n=52)
Low/No Expression (n=85)
Wilcoxon p=0.043*

MUC4-220
High Expression (n=31)
Low/No Expression (n=116)
Wilcoxon p=0.034*

MUC15-203
High Expression (n=50)
Low/No Expression (n=97)
Log-Rank p=0.030*
Wilcoxon p=0.030*

MUC16-201
High Expression (n=74)
Low/No Expression (n=73)
Wilcoxon p=0.031*
worse survival in high PC4 patients with an average survival time of 632 days compared to 804 days in low PC4 patients (p=0.2018, Fig 2.8F).

2.B.4. Identification of mucin splice variants associated with PDAC survival

The discordance of individual mucin expression associating with expected patient survival in TCGA samples, either with cellularity-uncorrected or cellularity-corrected values, led me to consider the expression of alternatively spliced mucins. Therefore, I realigned the raw RNA-Seq reads from TCGA patients to the current Ensembl annotations for all mucin transcripts. I generated survival plots for all mucin transcripts (n=110), excluding those expressed by less than 10 subjects (n=13), and calculated log-rank and Wilcoxon significance tests. I observed 12 variants associated with survival differences with higher expression of 5 SVs associated with better survival (Fig 2.9A) and higher expression of 4 SVs associated with worse survival (Fig 2.9C). Higher expression of four MUC1 transcripts was associated with statistically significant better survival (SV-210: 695 vs. 473 days, p=0.013; SV-221: 695 vs. 532 days, p=0.038; SV-222: 732 vs. 498 days, p=0.022; SV-226: 695 vs. 598 days, p=0.034), with MUC1-SV-210, -221, and -226 showing significant survival differences in early deaths (respective Wilcoxon p-value reported). Additionally, better survival outcome was observed in patients with higher expression of full-length MUC13 (SV-205: 695 vs. 498 days, p=0.015) and lower risk of early death in higher expression of MUC20 (SV-208: 691 vs. 593 days, p=0.044). Significant risk of earlier death was observed in patients with high expression of two MUC4 variants (SV-215: 593 vs. 634 days, p=0.043; SV-220: 460 vs. 652 days, p=0.034) and full-length MUC16 transcript
Additionally, higher expression of two other mucin transcripts was significantly associated with worse survival, MUC15 (SV-203: 486 vs. 684 days, p=0.030) and MUC21 (SV-201: 518 vs. 691 days, p=0.044), and a higher risk of early death in patients with higher expression of MUC22 (SV-201: 607 vs. 614 days, p=0.04). I assessed the expression level of these 12 transcripts (9 of which were robustly expressed across patients) (better survival SVs Fig 2.9B; worse survival SVs Fig 2.9D) and saw that nearly all subjects expressed the four MUC1 transcripts (75th quantile: SV-210=2.2 TPM; SV-221=5.13 TPM; SV-222=44.28 TPM; and SV-226=8.32 TPM). Just over 70% of patients expressed MUC4-SV-215 (75th quantile=1.54 TPM) and over 40% expressed MUC4-SV-220 (75th quantile=0.89). All patients expressed MUC13-SV-205 (median=337.1 TPM) and MUC20-SV-208 (median=9.07 TPM). All but one patient expressed MUC16-SV-201 (median=2.66 TPM). Over 67% of patient samples demonstrated expression of MUC15-SV-203 (75th quantile=1.63), while almost 81% had low expression of MUC21-SV-201 (75th quantile=1.04 TPM).

2.B.5. Determining PDAC Tumor Cellularity with PCR

Evaluation of tumor cellularity for samples stored on the Genomic Data Commons site of TCGA is assessed using the reported tumor purity score ascertained by the ABSOLUTE Purity Algorithm. However, analysis of gene expression in validation samples using PCR or simple transcriptomic sequencing required the innovative design of an alternative technique. By correlating all transcriptomic data available for all TCGA patients with the cellularity score of each respective sample, a list of genes highly associated with tumor purity was
Chapter II: Mucin Expression Profiling

calculated. The top 20 genes were plotted with their $R^2$ LogWorth and p-values (Fig 2.10A). The top gene hit was epithelial splicing regulatory factor 2 (ESRP2). However, in silico validation of this single gene lacked the appropriate level of correlation. Therefore, the top two genes were selected: ESRP2 and protein tyrosine kinase 6 (PTK6). The combination of both genes greatly strengthened the predictability of sample cellularity. Because these genes were to be assessed in PCR, the top negatively correlated gene, melanoma associated antigen H1 (MAGEH1), was also selected for screening. The advantage of use MAGEH1 in PCR validation of PDAC tumor sample cellularity was that it confirmed successful gene detection and counterweighted the detection of the positively associated genes. The cellularity of samples was qualitatively determined by taking the ratio of MAGEH1 to the average of ESRP2 and PTK6. Using a threshold of 1, high calculated samples were retained as ‘High Cellularity’ while low samples were excluded as ‘Low Cellularity’ (Fig 2.10B).

2.B.6. Experimental validation of MUC13 and MUC4Δ6 SV

I focused on two SV, one with negative impacts on survival (MUC4-SV-215), and one with a positive association with better overall survival (MUC13-SV-203), for the experimental validation. The MUC4-SV-215 demonstrated in-frame skipping of exon 6 (MUC4Δ6), which corresponds to the initial portion of the NIDO domain. MUC13-SV-203 codes for the full-length protein (MUC13WT). I pursued validation of MUC4Δ6 and MUC13WT in a separate patient tumor sample set obtained from Whipple resections from the UNMC Tissue Sciences Core Facility.
Figure 2.10: Genes Panel for PCR-based Cellularity Assessment

(A) Correlation of transcriptomes to sample cellularity was conducted and the top 20 most significant genes plotted by correlation LogWorth and p-value after false discovery rate (FDR) correction. ESRP2 and PTK6 were selected as the positively correlated genes and MAGEH1 was selected as the negative correlation gene. (B) Thresholding patient samples with the 3 gene panel demonstrated 3 samples were Low Cellularity while 15 were High Cellularity.
FIGURE 2.10: GENES PANEL FOR PCR-BASED CELULARITY ASSESSMENT

A

B

Chapter II: Mucin Expression Profiling
Figure 2.11: Validation of the expression of MUC13-sv205 and MUC4-sv215

In a separate patient tumor set (n=17) by ddPCR: (A) Of 26 candidate genes, three had the largest Spearman’s rho: two positively correlated, ESRP2 (rho= 0.6191, p≤0.001) and PTK6 (rho=0.7682, p≤0.001), and one negatively correlated, MAGEH1 (rho=-0.6570, p≤0.001). Based on the established use of CK19 as an epithelial marker, we included it in this analysis to compare its correlation to the other three (rho=0.4782, p≤0.001). (B) We used the ratio of the average of PTK6 (median expression= 22.8 copies/1000 GPI copies (c/rc)) and ESRP2 (median expression= 73.4 c/rc) to MAGEH1 (median expression= 19.1) expression. Each line and color correspond to a single patient. (C) Expression of GPI normalized MUC13 yielded a median of 56.1 c/rc. (D) Expression of GPI normalized MUC4WT (median expression= 31.6 c/rc) and MUC4Δ6 (median expression= 51.4 c/rc). Individual patient samples were plotted to appreciate the difference between both transcripts (lines and colors indicate the patient labels). Interestingly, all cases demonstrated higher expression of MUC4Δ6 than MUC4WT except for patients 2 and 6. (E) Stratifying the expression of MUC13WT above (n=8) or below (n=7) the median, higher expression trended toward better survival (758 vs. 445 days; p=0.1048). (F) MUC4Δ6 above (n=7) or below (n=8) the median, higher expression was associated with worse survival (393 vs. 801 days, p=0.0375).
FIGURE 2.11: VALIDATION OF THE EXPRESSION OF MUC13-SV205 AND MUC4-SV215

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESRP2</td>
<td>0.6243</td>
<td>0.001</td>
</tr>
<tr>
<td>MAGEH1</td>
<td>0.6312</td>
<td>0.001</td>
</tr>
<tr>
<td>PTK6</td>
<td>0.6994</td>
<td>0.001</td>
</tr>
<tr>
<td>CK19</td>
<td>0.4005</td>
<td>0.001</td>
</tr>
</tbody>
</table>

B

Gene Expression [Log Copies/1000 GPI Copies] vs. Mucin Expression [Copies/1000 GPI Copies]

C

Patient-based Mucin Expression [Log Copies/1000 GPI Copies]

D

Mucin Expression [Copies/1000 GPI Copies]

E

MUC13WT

High Expression (n=8)
Low Expression (n=7)
p=0.10

F

MUC4Δ6

High Expression (n=7)
Low Expression (n=8)
p=0.03

Overall Survival (days)
(n=17) by copy number analysis (CNA; digital PCR). To assess genes associated with cellularity, the transcriptome was correlated with the ABSOLUTE purity score. The top two positively correlated and the top negatively correlated genes, in addition to the canonical keratin-19, were selected for assessment in the validation set (Fig 2.11A). The malignant proportions of tumor samples were calculated based on the median expression of PTK6 (7.7 in the low cellularity and 22.8 copies/1000 copies GPI in the high cellularity), ESRP2 (29.4 in low cellularity and 94.8 copies/1000 copies GPI in high cellularity), and MAGEH1 (46.2 in the low cellularity and 24.7 copies/1000 copies GPI in the high cellularity) (Fig 2.11B). The sample cellularities were calculated and stratified into high (n=15) and low (n=2). Overall, expression of MUC13WT, MUC4Δ6, and MUC4WT was higher in the high vs. low cellularity. Comparing the high to low cellularity groups, the median expression of MUC13WT was 56.1 to 16.9 (Fig 2.11C), MUC4Δ6 was 49.1 to 13.2, and MUC4WT was 23.2 to 6.3 copies/1000 copies of GPI (Fig 2.11D). Due to a lack of a reliable cellularity normalization calculation in our CNA, the low cellularity samples were excluded from subsequent survival statistical analyses.

Patients were stratified about the median expression of MUC13WT and MUC4Δ6 into high (n=8 and 7, respectively) or low expressers (n=7 and 8, respectively), and Kaplan-Meier survival curves were plotted. Despite the low number of available samples, an appreciable trend in survival difference was observed in both SVs, which supported our in-silico findings. The overall median survival time was 451 days, but the difference was exaggerated when considering the stratification components. High expressers of MUC13 had a median survival
time of 758 days, whereas low expressers had a median survival of 445 days (p=0.1048, Fig 2.11E). High Expressers of MUC4Δ6 had a median survival time of 393 days compared to 801 days in the low expressers (p=0.0375, Fig 2.11F).

2.C Discussion

The expression and function of several mucins have been investigated within the context of PDAC; however, the scope of these studies is generally limited to single mucin. The establishment of cancer transcriptome databases has increased our ability to assess the potential role of previously unrecognized genes and their variants in specific malignancies. Other studies have reported employing widely used databases to conclude that the expression of MUC4, especially when combined with MUC16 and MUC20, was significantly associated with worse survival outcomes in PC patients. Until present, no study has addressed the need for cellularity-based correction of the RNA-Seq dataset due to the variable and often a low number of malignant cells present in PDAC tissues. This seemingly innocuous oversight obscures the real effects of genes expressed by cancer cells within PDAC tumors. Furthermore, albeit more detrimental, is the indiscriminate use of databases that fail to control for or stratify reported cases adequately. Here I ensured that PDAC samples were correctly vetted before selection. The difference in survival outcome in pancreatic cancer is intrinsic to the type, with PDAC have significantly worse survival. Indiscriminate stratification of TCGA samples by mucin expression essentially separates neuroendocrine tumors from PDAC, which the former having substantially better survival.
The concern of malignant cellularity was addressed by the TCGA Consortium, which used a computational algorithm (ABSOLUTE Scoring) to quantitatively determine the cancer cell population of tumor samples based on gene expression and chromatin methylation states. Based on these scores, they assigned qualitative cellularity labels (high or low cellularity) to each sample. This method, while resourceful, poses a statistical dilemma. While it permits more accurate gene expression analysis from patient to patient, it effectively halves the sample population and reduces statistical power. Because healthy pancreatic tissues are absent of nearly all mucins while PDAC cells express progressively increasing amounts, I elected to use the malignant purity scores of all samples to “normalize” the expression of mucins.

Using the mucin expression profiles, I explored a novel method for subtyping PDAC tumors to better understand different routes of disease progression, with the chief objective being better therapeutic approaches in the clinic and ultimately improved survival of patients. The most accepted subtyping strategies involve complicated and expensive sequencing of tumor samples to either identify pathways impacted by genomic mutations, or transcriptomic profiling of the whole tumor sample or microdissected samples. I have expanded on the well-established aberrant localization and overexpression of mucins to define disease subtyping based on genes correlated with four groups of co-expressed mucins. Pathway analysis of the correlated genes to each of the four mucin clusters suggests that mucin expression might signal or be involved in unique molecular fingerprints of PDAC tumors. I believe that the expression profile
of mucin in PC1 promotes the immunological reaction, as indicated by the many T-cell activation pathways. I have also shown that patients with a high expression of these mucins survive longer than patients with low expression. However, I am constrained by the incomplete data regarding lines of therapy given to the patients in the TCGA dataset. Therefore, more studies are necessary to substantiate our hypothesis.

Numerous studies have reported that malignant tumor cells demonstrate a wide array of abnormal alternative splicing events in their expressed genes, some of which may have novel or unregulated functions, prognostic implications, or diagnostic potential with clinical implications. Therefore, I employed a focused bioinformatics-based approach to investigate the expression of mucin splice variants in PDAC tumor samples from TCGA patients. Expression of four MUC1 transcripts revealed improved survival times, as did the expression of full-length MUC13 and MUC20, while both MUC4 and the MUC15, full-length MUC16, MUC21, and MUC22 SVs demonstrated decreased survival of PDAC patients. The mechanisms by which these transcripts contribute to changes in survival outcomes are not understood and require future study. Expression of MUC1 is typical of many tissues and cell types, including gastrointestinal epithelium, stromal cells, and immunocytes. Thus, the detection of these transcripts may indicate the presence or activation of cells that impede aggressive disease biology. MUC16, the largest described mucin, is not fully characterized, and its contributions to biological and clinical aspects of PDAC are hypothesized but not well documented. MUC4 is expressed in isolated tissues; however, its
expression has been well established in PDAC cell lines and patient tumor samples and absent in healthy pancreas. Other groups have demonstrated that expression of MUC13 is associated with a more aggressive PDAC phenotype in cell line models. However, our analysis from TCGA, as well as within our validation samples, contradicts these findings. This disparate observation may be explained by assessing the functional status of MUC13 during tumorigenesis.

Notwithstanding these observations, no studies have evaluated the tumor-specific role and diagnostic potential of mucin SVs in PDAC. Our investigation of mucin transcripts in PDAC transcriptomes demonstrated that splicing of exon six from MUC4 presented with significantly decreased patient survival. This exon codes for the N-terminal sequence of the NIDO domain. Despite an incomplete understanding of the NIDO functionality, it has been linked to interactions between the expressing cell and the surrounding extracellular matrix. Interruption of the NIDO domain may permit loose adherence of PDAC tumor cells and increase their mobility. However, I detected the expression of this SV concurrently with MUC4WT, suggesting that the interaction of MUC4Δ6 with MUC4WT may result in reduced patient survival. If true, this further suggests that tumor cells expressing this transcript possess an increased metastatic potential and elevate disease aggressiveness.

Further, I found that high expression of MUC4Δ6 is an adverse prognostic marker and presented with significantly shortened survival in our validation PDAC patients. In contrast, expression of MUC13WT was discovered to be a favorable
prognostic finding and presented with more prolonged survival. Our ability to observe statistically significant differences in survival from TCGA comes from its large sample size. It is worth noting that the overwhelming majority of PDAC patients are diagnosed with late-stage disease compared to the early-stage cases represented in TCGA. The availability of patient samples for our validation studies is significantly limited and may have contributed to the lack of statistical significance in survival time when considering MUC13WT expression. Our analysis is likely underpowered to detect differences in survival due to this limitation. Nonetheless, I maintain that expressions of MUC4Δ6 and MUC13WT are unique to PDAC and should be considered when assessing the outcome expectation of patients.

2.D Methods

2.D.1. Data Acquisition and Characterization

Our study was initialized by acquiring the normalized RNA-Seq data (aligned to Genome Reference Consortium Human Build 38) of PC subjects (PAAD) from TCGA along with the matched clinical details. Only confirmed PDAC cases with complete clinical and cellularity data were retained for analysis (n=149 PDAC, n=3 normal adjacent). The expression of cell type-specific genes was assessed in all cases for amylase 2A (acinar cells), CD45 (immune cells), cytokeratin 19 (epithelial cells), leptin (adipose), and mesothelin (stromal cells). Our secondary in silico cohort was comprised of the PACA-AU pancreatic cancer downloaded from the International Cancer Genome Consortium (ICGC) and
consisted of n=67 samples after filtering n=24 samples due to histological diseases other than PDAC. The data were converted from the FPKM normalized values reported in ICGC to TPM by the formula:

\[
TPM_i = \left( \frac{FPKM_i}{\Sigma_j FPKM} \right) \times 10^6
\]

in order to make intergenic, cross-samples, and cross-dataset comparisons.

2.D.2. Realignment of RNA-Seq. Fragments and SV Calling

To account for updates to genome annotations, raw RNA-Seq reads were acquired from the NIH Genomic Data Commons site. These reads were aligned to the Ensembl 94 GRCh38 cDNA reference transcriptome [RRID:SCR_002344] using Salmon \(^{206}\) and the standard 31 k-mer index for high mapping accuracy.

2.D.3. Sample Cellularity Correction of Mucin Expression

The majority of mucins are expressed only by malignant epithelial cells of PDAC. Therefore, the detected mucin mRNA transcript levels from each patient required consideration of the proportion of tumor cells within the sample. Fortuitously, each PDAC sample in the TCGA dataset reports the matched cellularity score. This value was computed by the TCGA Consortium using the
ABSOLUTE algorithm, which assesses PDAC-specific transcripts and comparative chromatin methylation states between PDAC and healthy pancreatic cells. The result of the calculation is reported as a single number between 0 (all normal cells) and 1 (all tumors cells). The normalized expression of all mucins was divided by the respective cellularity scores. This permitted the inclusion of all patients in a single population and comparable to each other rather than dividing the cases into qualitative “high” versus “low” cellularity.

2.D.4. Mucin Splice Variant Validation in Patient Samples

In silico results were verified in a separate patient cohort (n=17) using PDAC samples collected from pancreaticoduodenectomy procedures performed at The University of Nebraska Medical Center. Total RNA was isolated from frozen PDAC tissues by using the mirVana miRNA Isolation kit (Ambion). Complementary DNA (cDNA) was synthesized using the iScript RT Supermix kit (Bio-Rad). Primers were designed to amplify MUC4WT, MUC4Δ6, both MUC4WT and MUC4Δ6 (MUC4-dual, for traditional PCR techniques), and MUC13WT. Primers designed for all are reported in Table 2. Glucose phosphate isomerase (GPI) was used as the reference gene and selected primers are reported in Table 3. Cellularity correlated genes were assessed in the TCGA PDAC dataset to determine a plausible method to determine cellularity in the validation samples by PCR-based techniques. Based on the analysis of cellularity-correlating genes, two positively correlated genes, protein tyrosine kinase 6 (PTK6) and epithelial splicing regulatory protein 2 (ESRP2), and one negatively correlated gene, melanoma-
<table>
<thead>
<tr>
<th>Target</th>
<th>Direction</th>
<th>Sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC4WT</td>
<td>Forward</td>
<td>AAAGGTCACGTGGGTCAATGC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCTATGCCCTGT TTCTCTACCAGA</td>
</tr>
<tr>
<td>MUC4Δ6</td>
<td>Forward</td>
<td>GGGACCACATTTTATCACAGAAGCAACACCTACC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACCGAGGCTGGCTTTTCAGCCACT</td>
</tr>
<tr>
<td>MUC4-dual</td>
<td>Forward</td>
<td>TCATCTTCCCAGAGTCAGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTAGAGAAACAGGGCATAGG</td>
</tr>
<tr>
<td>MUC13</td>
<td>Forward</td>
<td>TCCAGTCTCAAGTGTCCTGATG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTCGAGTCCACTGTAGCCAA</td>
</tr>
<tr>
<td>Target</td>
<td>Direction</td>
<td>Sequence (5’ -&gt; 3’)</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>GPI</td>
<td>Forward:</td>
<td>CAGACCCAGCACCCCCATAC</td>
</tr>
<tr>
<td></td>
<td>Reverse:</td>
<td>CCAAGGCTCCAAGCATGAATG</td>
</tr>
</tbody>
</table>
associated antigen H1 (MAGEH1), expression were evaluated using the primers in Table 4.

Gene expression was assessed using digital droplet PCR (Bio-Rad QX200 AutoDG system) and quantifying absolute transcript copies from approximately 12,000-20,000 droplet reads per target in each sample. Copy numbers were calculated within QuantSoft Analysis Pro (Bio-Rad Laboratories, RRID:SCR_008426) and back calculated to the copies per µL of input total RNA isolated from patient tumor samples. Copy numbers of all genes were normalized using respective GPI copies and reported as the number of transcripts per 1,000 copies of reference. Malignant cellularity of all samples was determined by dividing the average normalized copies of PTK6 and ESRP2 by MAGEH1 (Equation 2).

**EQUATION 2: CALCULATION FOR PDAC SAMPLE CELLULARITY BY PCR**

\[
\text{Cellularity} = \frac{\left(\frac{\text{PTK6} + \text{ESRP2}}{2}\right)}{\text{MAGEH1}}
\]

Score ratios >1 were labeled as “High Cellularity” and all <1 were labeled “Low Cellularity”. Expression of MUC13WT and MUC4Δ6 were stratified about the median c/cr and the overall survival.
<table>
<thead>
<tr>
<th>Target</th>
<th>Direction</th>
<th>Sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTK6</td>
<td>Forward</td>
<td>TCGGAACCGTGGTTCTTTGG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CGCCGCCAGATCTTGTAGT</td>
</tr>
<tr>
<td>ESRP2</td>
<td>Forward</td>
<td>GACAGATTAAGGATACTGTTGACCT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTAAAGGAACTGGTTCTCAGTTATG</td>
</tr>
<tr>
<td>CK19</td>
<td>Forward</td>
<td>TAGTGAGCGGCAGAATCAGG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATCTTCCTGTCCCTCGAGCA</td>
</tr>
<tr>
<td>MAGEH1</td>
<td>Forward</td>
<td>GAGTTTGTGCAGAGGGTA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CATCGTCCGAATCCAGTCA</td>
</tr>
</tbody>
</table>
2.D.5. Statistical Analysis

Mann-Whitney-Wilcoxon (for 2 groups) or Kruskal-Wallis (for 3 groups) tests were used to compare mucins between groups before and after cellularity-correction. Post-hoc Dunn’s tests were calculated, followed by Bonferroni corrections to adjust pairwise comparisons between groups. Significance levels within sample types, cellularity, and cellularity correction state (corrected vs. uncorrected) groups were corrected for false discovery errors in multiple comparisons using the Benjamini-Hochberg method. Overall survival time and cause of death were obtained and matched to respective patients from the supplemental clinical data available from the National Institute of Health’s GDC data portal or UNMC/Nebraska Medicine EHR. Survival time was measured in days starting at diagnosis and ending on the patient’s death or end of the follow-up period. Kaplan-Meier method was used to estimate overall survival distributions by mucin expression, categorized at the zero-excluding median as “high” or “low/no” expression. The log-rank and Wilcoxon tests were used to compare survival distributions. Principal component analysis by Spearman’s correlations was calculated for all subjects using cellularity-corrected expression followed by factor analysis using orthogonal rotation, with no significant differences between unrotated and rotated eigenvectors. Based on the scores of each subjects’ PC, they were stratified to the respective PC with top 25% (High), bottom 25% (Low), and middle 50% (Moderate). Pathway analysis was conducted using Qiagen Ingenuity Pathway Analysis software (RRID:SCR_008653) based on the fold-change of significantly correlated genes between high and low PC patients.
Prediction of all patients to PCA-based subtypes was calculated by strongly independent (naïve) Bayes analysis using 70% of the patients for training and 30% for validation by receiver operating characteristic curves for the validation samples. All statistical analyses conducted here were completed using JMP Pro statistical suite or R.

2.D.6. Power Analysis

In our validation group, there were n=15 retained patient tumor samples; and selecting an alpha=0.05, power=0.8, and a standard deviation=150, the analysis is powered to detect a difference of 234.8.

2.D.7. Sex as a Biological Variable

The incidence of pancreatic cancer in males versus females is not statistically different. Within TCGA PDAC samples, n=69 are female and n=81 are male. Therefore, the sex of patients was not applicable to the design or conduct of this study and was not considered.

2.D.8. Institutional Review Board

Data obtained from TCGA was approved following raw data acquisition request submission and was received de-identified. All data were downloaded, stored, and processed following the requirements established by the National Cancer Institute’s Genomic Data Commons. Patient samples used in our study were submitted to the UNMC tumor bank following Whipple resections and consenting to have surplus tissue used for research purposes. Samples were
procured by our lab under Aim 1 of IRB Protocol # 186-14-EP. Samples were blinded by central processing upon intake to the tissue bank. This study was computational and confirmatory in nature, and as such, the randomization of subjects was not necessary.
CHAPTER 3:

FUNCTIONAL ANALYSIS OF MUC4Δ6
3.A Introduction

In summary of Chapter One, a bioinformatics analyses of TCGA revealed, in part, that PDAC patients expressing the alternatively spliced mucin 4 (MUC4) variant have significantly reduced overall survival. While aberrant expression of many mucins is a well described phenomenon in PDAC, the exact involvement of mucins in disease pathobiology is complex and foundational understanding is incomplete. Some studies investigating the role of MUC4 in PDAC aggressiveness have suggested it may be involved in tumorigenesis, appearing early in the transformative process, tumor cell migration, and therapeutic resistance.

MUC4 is a large membrane-bound glycopeptide composed of 12 functional domains coded by 26 exons. The exact role of each domain in MUC4 is poorly understood but they have been suggested to contribute to matrix and adjacent cell interactions and intracellular signaling. When considering the implication of altered domain coding through alternative slicing, a few studies have demonstrated at least 22 different transcripts of MUC4. The variant identified in Chapter One of this work contains an in-frame deletion of exon 6 (termed MUC4Δ6). Exon 6 is a 156 base pair sequence coding for 52 amino acids at the N-terminus of the nidogen-1-like (NIDO) domain downstream of the substantial tandem repeat region which lies upstream of the adhesion-associated (AMOP) domain. I hypothesized that splicing of this exon induces a steric hinderance between the extracellular matrix and the full NIDO and AMOP domain, effectively reducing the binding potential or strength of MUC4 to
MUC4 protein is divided into 2 principal segments about its auto-cleavage site (GDPH) generating a freed extracellular MUC4α, which enables MUC4 to interact with surrounding stroma and adjacent cells, and membrane-tethered MUC4β, which includes three suspected EGF-like domains and a von Willebrand domain (vWD). MUC4α contains the hyper-glycosylated extensive tandem repeat domain, a NIDO-like domain, and an AMOP domain. MUC4Δ6 demonstrates an in-frame deletion of exon 6, which codes the 52 most N-terminal amino acids of the NIDO domain adjacent to the tandem repeat.
FIGURE 3.1: DOMAIN MOIETY OF MUC4
the matrix. If true, the consequence of this modification would be a molecule with reduced adhesion properties and an increase in migratory potential of the expressing cell. To test this hypothesis, cell lines were established to enhance the expression of MUC4Δ6 selectively.

3.B Results

The exploration of the functional role of MUC4Δ6 required the generation of novel cell lines with enhanced expression of the MUC4 splice variant. Because the MUC4Δ6 is expressed concurrently with MUC4WT, a cell line expressing the sole MUC4 variant would fail to recapitulate the actual tumor environment more accurately. Therefore, attempts were made to offset the ratio of MUC4WT to MUC4Δ6 expression.

3.B.1. MUC4WT Knockdown

In line with the hypothesis that MUC4Δ6 functions in concert with MUC4WT on the expressing cell, the first attempt to increase the splice variant was conduct using transfection of shRNA targeting exon 6 of MUC4. In this way, wild-type was expected to decrease and the consequential upregulation of MUC4 expression would increase the overall abundance of MUC4Δ6 in natively expressing cells. Transfection of three shRNAs (see Methods and Materials) was conducted in CD18/HPAF-II, T3M4, Capan-1, and COLO357 pancreatic cancer cell lines. This knockdown failed in all cell lines except COLO357. Semi-quantitative PCR using the dual MUC4 primers described in Methods and Materials revealed shMUC4E6 #2 had the maximum reduction in MUC4WT
expression with a drop of about 15% and an almost 25% increase in MUC4Δ6 expression (Fig 3.2A and 3.2B). However, these changes were not substantial enough for consideration moving into functional studies. Further, the use of puromycin antibiotic, necessary to stabilize and maintain enriched populations of transfected cells, induced expression of MUC4 with vast amounts of wild-type transcript (Fig 3.2C). As a result, this method of splice variant overexpression was abandoned.

3.B.2. miniMUC4d6 Overexpression

Following the unacceptable modulation in MUC4Δ6 to MUC4WT expression in my shRNA knockdown cell lines, I decided to attempt to establish an overexpression system. I began with a miniature MUC4 construct containing 10% of the tandem repeat domain sequence of wild-type MUC4 (miniMUC4) and generated a miniMUC4Δ6 overexpressing construct (described in Methods and Materials). The size of the MUC4 gene makes molecular alterations and sustained fidelity of sequence quite challenging. Several PST-miniMUC4Δ6 transformed bacterial colonies were confirmed by selective restriction digestions and gel resolution, as well as DNA sequencing. CD18/HPAF-II, COLO357, and MiaPaCa2 cell lines were transfected and screened. By immunoblot detection, no transfected cultures were confirmed positive for expression of the target gene using an anti-MUC4 antibody or an anti-c-Myc-tag antibody, the epitope present on the C-terminus of vector coded transgene. RNA was isolated from transfected CD-18/HPAF-II and COLO357 cell lines. Complementary DNA was synthesized from
Figure 3.2: MUC4Δ6 Expression Enrichment by shRNA Knockdown of Wild-Type

(A) Transfection of shRNA virus into COLO357 PDAC cells demonstrated limited reduction in MUC4WT expression or increase in MUC4Δ6 expression. (B) Gel resolution of dual MUC4 PCR amplicons. (C) Treatment of cells with puromycin selection antibiotic enhanced overall expression of MUC4.
FIGURE 3.2: MUC4 Δ 6 EXPRESSION ENRICHMENT BY SHRNA KNOCKDOWN OF WILD-TYPE
isolated RNA and primers from the exon 6 flanking regions were used to amplify the presence of PST-miniMUC4Δ6 vector sequence in transfected COLO357 cells. Gel resolution of this confirmation PCR revealed that the cells contained and were expressing the plasmid transgene with the appropriate size (Fig 3.3).

3.C Discussion

My results from Chapter 2 of this dissertation demonstrated that pancreatic ductal adenocarcinoma patients expressing high alternative splice variant MUC4Δ6 had significantly reduced median overall survival outcomes. These observations were in accord by in silico analysis and by PCR in validation Whipple samples. Previous studies have suggested that the NIDO domain of MUC4 enables the expressing cell to adhere tightly to neighboring cells and surrounding stromal tissue. Additionally, some limited studies have postulated that NIDO further enables MUC4 to interact with neural sheaths. This hypothesis is exceedingly relevant because PDAC frequently demonstrates perineural invasion. Based on these limited in vitro and computational studies, I hypothesized that MUC4Δ6, in concert with MUC4WT, might reduce the overall binding of expressing cells to the ECM and increase interaction with myelin cells to facilitate an aggressive nerve fiber invasion. To begin to uncover the possible mechanisms behind these observations, a model system of MUC4Δ6 was required. At first attempt, I designed an shRNA targeting exon 6 of MUC4 and transfected several PDAC cell lines. Though some relative enhancement of MUC4Δ6 expression was observed, the overall modifications were subpar for their inclusion in experimental investigations. The next concept attempted to overexpress MUC4Δ6 in PDAC cells
COLO357 cells transfected with the new miniMUC4Δ6 overexpression vector were assessed for expression of transgene by PCR. The non-transfected control cells demonstrated the proper amplicon size matching miniMUC4 since the miniMUC4 amplicon contained the same sequence as full-length MUC4. All four miniMUC4Δ6 colony transfected cells demonstrated expression of the intended truncation amplicon. To match the sequence of the insert in the miniMUC4 pST vector, it was included as a control reaction.
Chapter III: Functional MUC4Δ6

FIGURE 3.3: MINIMUC4Δ6 OVEREXPRESSION PCR IN COLO357 CELLS
natively expressing MUC4WT. I modified a miniature MUC4 construct previously designed by our laboratory that coded a 90% reduction in the tandem repeat domain to exclude exon 6. Despite confirmation of proper molecular techniques and results, as well as sequencing and PCR validations, I was never able to establish the overexpression cell line. This failure was consistent over nine independent attempted transformations and four reattempts at construct synthesis.

The major hurdle with the induction of a MUC4 gene is clearly its size. Full-length MUC4 RNA is almost 16.8 kilobases coding a transmembrane core protein that is around 230 kDa and as large as 900 kDa with its full glycocalyx. Molecular cloning of full-length MUC4 is unlikely to be successful due to this prohibitive mass. Modification and cellular manipulation of MUC4 is challenging to say the least. Given that transfection of my miniMUC4Δ6 was unsuccessful, other potential modalities of selective expression have been conceptualized. One very optimistic idea to generate MUC4Δ6 expressing cell lines would involve the use of a CRISPR/Cas system. The major drawback to this approach is gene edited cells would exclusively express the splice variant without the ability to express wild-type MUC4. Alternatively, a MUC4Δ6 cassette could plausibly be inserted into the genome of cells. A potentially better approach would involve the use of small interfering oligonucleotides to block the splice sites of exon 6 and promote the selective splicing of the gene. This approach has merit and is well documented. Studies to correct the mutation in the dystrophin protein in muscular dystrophy has successfully induced the splicing exclusion of exon 51 in the mutant transcript leading to a truncated but functional protein. Several molecules carrying the
splicing directive oligonucleotide have moved into human clinical trials and the results are exceedingly promising. Nonetheless, my bioinformatics and validation results suggest MUC4Δ6 is an important aspect of PDAC pathobiology, and the continued investigation into this splice variant seems obvious.

3.D Methods


Three shRNA oligonucleotide pairs were designed to target exon 6 of full-length MUC4 (Table 5) with Bgl-II and Hind-III restriction site sequences matching the cloning restriction sites of the pSUPER.retro.puro (pSRP) short interfering RNA expression vector (OligoEngine VEC-PRT-0002). Each oligonucleotide was dissolved to a working concentration of 3 mg/mL in molecular grade water. To anneal oligonucleotide pairs for insertion to the vector backbone, 1 µL of each was combined in a 0.2 mL microcentrifuge tube along with 48 µL of annealing buffer (100 mM NaCl, 50 mM HEPES pH 7.4). The reaction was annealed on a thermocycler using the protocol: 90°C for 4 min, 70°C for 10 min, then step cooled 0.2°C per 30 sec to a final temperature of 10°C. The anneal insert was stored at 4°C until final plasmid preparation. The empty pSRP vector was linearized by adding 1 µg to a 0.2 mL microcentrifuge tube with 5 µL of buffer 3.1, 1 µL of Bgl-II restriction enzyme (New England BioLabs R0144S), and 1 µL of calf intestinal alkaline phosphatase (CIP, New England BioLabs M0525) in a final reaction volume of 50 µL. The reaction incubated at 37°C on a glass bead bath for 2 hours before cleaning and reconcentration using a DNA cleanup kit. The entire eluate
**TABLE 5: SHRNA AGAINST MUC4WT**

<table>
<thead>
<tr>
<th>Pair</th>
<th>Strand</th>
<th>Aligned Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>shRNA-1</td>
<td>Top</td>
<td>5' GATCCCGACGAGACGTTCATGGTGATICAAAGAGATCACTATAGAAGCTCTGTCTTTTA 3'</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>3' GGATTGCTCTGCAAGATACCACTTAAGTTCAGTAGTGATCTTTGAGAGCAAATAATTCGA 5'</td>
</tr>
<tr>
<td>shRNA-2</td>
<td>Top</td>
<td>5' GATCCCGTAAGGTCAGTGAGCTGCAATTCAAGAGATTGACCCGATGACCTTTTA 3'</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>3' GGATTTCCAAGTGCACCCAGTAAAGTTCCTAAACGTGGGTGCACTGGAAATATTTATTCGA 5'</td>
</tr>
<tr>
<td>shRNA-3</td>
<td>Top</td>
<td>5' GATCCCGAGATGACAAAACACGGGGTTCAGAGAGACCCCGTTCCTTTGTCATCTTTTTTA 3'</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>3' GGGCTCTACTGTTGTGCTCCCAAGTTCCTGGGGCAACACAGTAGAATAATTCGA 5'</td>
</tr>
</tbody>
</table>

Other: (BglII) (sense MUC4E6) (HindIII) (antisense MUC4E6)
was added to a 0.2 mL microcentrifuge tube with 5 µL of CutSmart buffer, 1 µL of Hind-III High Fidelity restriction enzyme (New England BioLabs R0104S), and 1 µL of CIP. The reaction was incubated at 37°C on a glass bead bath for 1 hr then gel resolved to validate proper linearization and isolate only dual cut vector free of the stuffer sequence. The annealed oligonucleotides were digested with both enzymes in the same manner described and gel purified. In a 3:1 w/w ratio, double digested insert and linearized vector backbone were mixed in a 0.2 mL microcentrifuge tube along with 2 µL ligation buffer and 1 µL T4 DNA ligase enzyme (New England BioLabs M0202S) in a total 20 µL reaction. The ligation reaction incubated at room temperature overnight. Competent E. coli bacteria were transformed as described in section 3.B.3 and individual colonies were screened for proper shRNA insertion. Colonies with confirmation by restriction digestion were sequenced to obtain final validation of molecular reactions.

Positive colonies were expanded, and target plasmids were isolated by midi-preparation. Phoenix-AMPHO cells (ATCC CRL-3213), which contain and express the amphotropic envelope protein transgene, were cultured, and expanded to 10 cm plastic cell culture dishes. Once cells were 50-75% confluent, cultures were serum starved for 6 hours. In a 1 mL microcentrifuge tube, 500 µL of serum-free media was added to 5 µg of plasmid, while in a second microcentrifuge tube, 500 µL of media was carefully added to Polybrene transfection reagent (1,5-Dimethyl-1,5-diazaundecamethylene polymethobromide, Sigma Aldrich TR-1003G) in a final concentration of 10 µg/mL. The tubes were flicked and briefly spun down to collect all solution. The plasmid mixture was
carefully added to the Polybrene mixture dropwise. The mixture was flicked several times, concentrated to the bottom of the tube, and incubated for 15 min. Immediately before transfection of pSRP-shRNA plasmid, cultures were replaced with 3 mL media. The plasmid-Polybrene mixture was added dropwise to cultures, and plates were incubated at 37°C overnight before replacing transfection with 10 mL of free complete culture medium. Every 24 hr for 5 days, all media was collected and replaced. Collected media was filtered through a 0.2 µm syringe filter into a 15 mL tube containing 5 mL of sterile filtered polyethylene glycol 6000 solution and mixed to precipitate lentiviral particles. The tubes were incubated at 4°C until 24 hr after the 5th media harvest. All tubes were centrifuged at 1500 x g, 4°C for 30 min. The supernatant was aspirated, and the concentrated lentivirus was combined from all five tubes, diluted to 500 µL with sterile PBS, and stored at -80°C until infection of target cells.

3.D.2. Lentiviral Infection of Target PDAC Cells

Target PDAC cell lines were cultured until approximately 50% confluency in a 10 cm plastic cell culture dish. The day of viral transfection, culture media was replaced with 3 mL serum-free media and serum starved for 6 hr. Cultures were inoculated with a viral titer from the previously prepared shRNA lentivirus and incubated at 37°C for 12 hr before the addition of 3 mL complete culture media. Transfected cell selection began 24 hr after lentiviral infection. Complete cell culture media containing 5 µg puromycin per mL media was added to cells and changed daily until all non-transfected control cells were dead. Cells cultures were
dropped to 1 µg puromycin per mL media for maintenance. When stabilized cells recovered and expanded, they were passaged and evaluated for target gene knockdown.

3.D.3. Overexpression of miniMUC4

Cloning of full MUC4 into mammalian cells is challenging due to its incredible size. Further, amplification of its tandem repeat domain is difficult due to its repetitive sequence and the high G-C content. Our group previously engineered a miniature construct of the MUC4 sequence that contains 10% of the tandem repeat. This construct was cloned into the pSecTag2C plasmid vector. I used this vector as the starting point for the engineering of our MUC4Δ6 transcript of interest (Fig 3.4). Sequence analysis of the cloned miniMUC4 sequence identified two single-cut restriction sites flanking the exon 6 sequence. PCR primers were designed to anneal upstream of the BspE1 restriction sequence with a reverse primer annealing at the base 5’ of the exon 6 sequence (5’ fragment). Immediately downstream of exon 6, a forward primer annealing to the exon 7 sequence was paired with a reverse primer annealing within the von Willebrand Factor (vWF) domain downstream of the Blp1 restriction sequence (3’ fragment). The primers (presented below in Table 6) adjacent to exon 6 contained the sequence coding for the Sfi1 restriction site.

A 25 µL PCR was conducted using 1µL pSecTag2C miniMUC4 (~3µg template), 67 mM Tris HCl; pH 8.3, 5 mM MgCl2 (Roche 10x PCR Buffer), 200µM deoxynucleotide triphosphate (dNTP, Roche DNTP), 0.2 µM each of forward and reverse primers, and 2 U of Q5 High Fidelity DNA polymerase (New England
Beginning with an in-house engineered miniMUC4 (coding for 10% of the tandem repeat domain) cloned into overexpression vector pSecTag C, fragments flanking exon 6 were amplified by PCR introducing a novel Sfi-I restriction site used to ligate the two fragments. The new Δ6 fragment and the parental pSecTagC-miniMUC4 are sequentially double digested with BspE-I and Blp-I before ligation of the modified sequence into the linearized partial miniMUC4 vector. The final product was amplified by transformation of a competent bacterial host and isolated by midi-preparation.
FIGURE 3.4: PSECTAG C-MINIMUC4Δ6 ENGINEERING

PCR exon6-flanking fragment

Digest at novel Sfi-I site (*)
Introduces Ala x5

Ligate new fragment into original vector backbone at Bpl and BspEl restriction sites
BioLabs M0491). Because the 5’ fragment was within the difficult to amplify tandem repeat, 2% dimethyl sulfoxide (DMSO, Sigma Aldrich D2650) was included. The 3’ fragment was amplified using a PCR program of 95°C for 3 minutes (min), 45 cycles of (95°C for 30 seconds (sec), 60°C for 30 sec, 72°C for 1min), 72°C for 5 min. The 5’ fragment was amplified using a touchdown PCR program of 95°C for 3 min, 45 cycles of (95°C for 30 sec, 60°C for 30 sec [decreasing 0.2°C each cycle], 72°C for 2 min), 72°C for 5 min. Amplicons were resolved by electrophoresis on a 2% agarose gel containing ethidium bromide (Fig 3.5A). Bands were careful extracted and purified amplicon fragments were isolated using a DNA gel extraction kit (New England BioLabs Monarch T1020). The amplicons were further amplified through a second round of PCR and purified using a PCR cleanup protocol (New England BioLabs T1030). To prepare the two fragments for re-ligation into the vector, the introduced Sfi1 restriction site was digested using 500 ng of amplicon, 5 µL of 10x CutSmart Buffer (New England BioLabs B7204), and 20 U of Sfi1 enzyme (New England BioLabs R0123) in a total reaction volume of 50 µL. The reaction was incubated in a glass bead bath at 50°C for 1 hour (hr) before purification with a PCR cleanup protocol. The undigested fragments and the ligated product were resolved and extracted from a 2% agarose/ethidium bromide gel as previously described. To confirm proper annealing, the ligated fragment was gel resolved (Fig 3.5B) and sequenced. The ligated amplicon was further amplified by 40 cycles of PCR using the most distal primers and Q5 high fidelity DNA polymerase and purified using a PCR cleanup. To ensure proper orientation and retainment of the Sfi1 site, the ligated amplicon was digested again
### TABLE 6: PSECTAG C- MINIMUC4Δ6 PRIMERS

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-Forward</td>
<td>ATATATTCCGGAATTCTGCAGATATCTCTGA</td>
</tr>
<tr>
<td>5’-Reverse</td>
<td>TTGGCCGCGGCCGGCCGCTGATAAAATGTGGTCC</td>
</tr>
<tr>
<td></td>
<td>CCCGA</td>
</tr>
<tr>
<td>3’-Forward</td>
<td>AAGCGGCCGCGCCGGCCAGCAACACCTACCAAG</td>
</tr>
<tr>
<td></td>
<td>CCATCC</td>
</tr>
<tr>
<td>3’-Reverse</td>
<td>TGGAGCGGTACTGAGCCG</td>
</tr>
</tbody>
</table>

*Introduced Sfi1 Sequence*
with Sfi1 and gel resolved (Fig 3.5C). The original pSecTag2C miniMUC4 plasmid was cut using the fragment restriction sites BspE1 (New England BioLabs R0540) at the 5’ end and Blp1 (New England BioLabs R0585) at the 3’ end using the same protocol described for Sfi1 but including 5 U of calf intestinal alkaline phosphatase (CIP, New England BioLabs M0525) and using a PCR cleanup between each enzyme. Concurrently, the ligated fragment was also dual digested in the same manner and purified using PCR cleanup. Linearized vector and clonal fragment were combined in a 1:3 volume ratio along with 2 µL of 10x T4 DNA Ligase buffer and 400 U of T4 DNA Ligase (New England BioLabs M0202) in a total reaction volume of 20 µL. The reaction was mixed and incubated at 25°C overnight.


After confirming ligation of exon 6 flanking fragments by gel electrophoresis and DNA sequencing (UNMC Genomics Core), the plasmid was used to transform high efficiency, endA1 deficient, competent E. coli (New England BioLabs 10-beta C3019). A vial of 10-beta cells was thawed on ice for 20 min before aliquoting 20 µL into a microcentrifuge tube. The entirety of the ligation reaction was added to the tube with bacterial cells. The tube was carefully flicked several times then incubated on ice for 30 min. The tube was then heat shocked at exactly 42°C for exactly 30 sec then placed back on ice for 5 min. In a 10 cm round culture dish, 10 mL of Luria-Bertani broth (LB) based agar (5 g peptone, 2.5 g filtered yeast extract, 2.5 g sodium chloride, 5 g agar in 500 mL and autoclaved) containing ampicillin was solidified and cooled to room temperature. Following the
final ice incubation, the bacterial cells were transferred and spread across the surface. The plate was covered and incubated upside down at 37°C, 5% CO2 overnight. The next day, colonies were individually picked and expanded in 5 mL LB broth overnight in a shaker incubator at 37°C and 200 rpm. Plasmids were isolated from 1 mL culture using a plasmid mini prep kit (New England BioLabs Monarch T1010). Plasmids from each colony were double digested with BspE1 and Blp1, as previously described, and resolved by electrophoresis (Fig 3.5D). Plasmids producing the expected fragment size were sent for DNA sequencing to confirm framing and proper sequencing.

3.D.5. Target Cell Transfection

Target mammalian cells were cultured in a 6-well plate from 5*10^6 until 90-95% confluent. Cultures were serum starved for 6 hr prior to transfection. In two microcentrifuge tubes, 100 µL of serum-free DMEM was added. In one tube, 10 µL of transfection reagent (TurboFect, Thermo Fisher R0531) was added while the other tube received 1-5 µg of plasmid DNA. The tubes were gently flicked and incubated for 5 min at room temperature. Then, the DNA mix was added into the transfection reagent dropwise. The tube was gently flicked several times and incubated at room temperature for 20 min. The media on the cells was aspirated and replaced with 1 mL serum-free DMEM before adding the transfection reagent-DNA mixture dropwise. The plate was gently swirled and incubated overnight at 37°C, 5% CO2. After 24 hr, 800 µL of DMEM with 10% bovine serum was added to the well. The media was completely aspirated and replaced with 2 mL DMEM
Figure 3.5: Restriction Digestion and Ligation of pST-miniMUC4Δ6

Each step of the modification of the miniMUC4 viral plasmid was visualized by agarose gel electrophoresis and confirmed by sequencing. (A) PCR of the exon 6 flanking fragments confirmed a proper and specific 514 bp 5' amplicon and 852 bp 3' amplicon. (B) Ligation of the two fragments following Sfi-I restriction digestion yielded a single amplicon of 1352 bp. (C) Validation of ligated clonal fragment re-digested with Sfi-I resulted in two fragments matching the original 5' and 3' exon 6 flanking amplicons. (D) Following transformation of modified pSecTagC miniMUC4Δ6 in competent bacteria, single colonies were picked from ampicillin LB agar plates and expanded. Plasmids were isolated and double restriction digested with BspE-I and Blp-I to confirm proper plasmid size. Colonies were compared to double digest pSecTagC miniMUC4.
Figure 3.5: Restriction Digestion and Ligation of pST-miniMUC4Δ6

A. Exon 6 Flanking PCR

B. Amplification and Ligation

C. Ligated Digestion Confirmation

D. Colony Validation Restriction Digestion
with 10% bovine serum after an additional 24 hr. After 48-72 hr, cells were exposed to media containing Zeocin selection antibiotic. After confirming miniMUC4d6 expression by PCR and western blot using anti-MUC4 clone 8G7 antibody, cells proceeded to in vitro experiments.

3.D.6. Western Blot Validation

Cells were grown in a 6-well plate until about 90% confluent. Wells were washed twice with sterile PBS buffer and lysed with 100-150 µL of fresh, cold radio-immunoprecipitation assay buffer (RIPA, 50 mm Tris-HCl, pH 7.4; 0.25% Na-deoxycholate; 1 mm EDTA; 150 mm NaCl; 1% NP-40) containing with 5 mM sodium orthovanadate, 5 mM sodium fluoride, and 1 mM phenylmethylsulphonyl fluoride. Wells were scraped and added to microcentrifuge tubes, exposed to a -80°C freeze-thaw cycle, vortexed, and passed through an 18G hypodermic needle. Lysates were centrifuged at 12,000 x g, 4°C for 20 min before supernatant was isolated to a new microcentrifuge tube and stored at -20 °C until gel electrophoresis. Lysates concentrations were standardized to a uniform 2 mg/mL, after BCA assay quantification, diluting in complete RIPA buffer and adding 6x β-mercaptoethanol protein dye. Prepared lysates were ran on 2% Tris-agarose gels for 4 hr at 120 V. Resolved proteins were transferred to PVDF membranes by wet passive diffusion overnight. Membranes were washed briefly with PBS and blocked with 5% fat-free milk in PBS for 1 hr, then washed again with PBS. Membranes were incubated with target (1:1000 anti-MUC4 8G7, 1:500 anti-MUC4 6E8, 1:2500 anti-β-actin, or 1:2000 anti-c-Myc 9E10) mouse anti-human
antibodies. Membranes were wash three time with PBS-0.1% tween-20 (PBS-T) buffer for 15 min each. Goat anti-mouse horseradish peroxidase secondary antibody (1:5000, Thermo Fisher 31420) for 1 hr. Following secondary incubation, membranes were again wash three times with PBS-T for 15 min each before 2 min incubation with horseradish peroxidase chemiluminescence reagent (SuperSignal West Dura, Thermo Fisher 34075). Luminescent membranes were exposed to reactive film and developed for visualization.

3.D.7. PCR Validation

RNA isolated from transfected cells was converted to complementary DNA synthesized using the iScript RT Supermix kit (Bio-Rad). A 25 µL PCR was conducted using 1 µL cDNA, 67 mM Tris HCl; pH 8.3, 5 mM MgCl2 (Roche 10x PCR Buffer), 200µM deoxynucleotide triphosphate (dNTP, Roche DNTP), 0.2 µM each of forward and reverse primers, and 2 U of Q5 High Fidelity DNA polymerase (New England BioLabs M0491). Amplicons were amplified using a PCR program of 95°C for 3 minutes (min), 40 cycles of (95°C for 30 seconds (sec), 60°C for 30 sec, 72°C for 1min), 72°C for 5 min. Amplicons were resolved by electrophoresis on a 2% agarose gel containing ethidium bromide and visualized with a gel UV dock (Bio-Rad) or careful extracted and purified using a DNA gel extraction kit (New England BioLabs Monarch T1020).
CHAPTER 4:

NANOPARTICLE-BASED CANCER-SPECIFIC

ALTERNATIVE SPLICE VARIANT DETECTION ASSAY
Chapter IV: AuNP Detection Assay

4.A Introduction

With the majority of PDAC patients diagnosed with advanced disease and the limitations of current PDAC therapies, the improvement of patient outcomes depends on early diagnosis. Towards this goal, substantial progress in PDAC diagnostic approaches and technologies is critical. Furthermore, the discovery of disease-specific biomarkers is necessary for the development of early detection assays. Currently, carbohydrate antigen-19-9 (CA19-9) is the only biomarker approved for PC. Clinicians use CA19-9 to monitor therapeutic response and detect disease progression. This marker, however, is not acceptable for diagnostic applications. Therefore, novel markers for the early detection of PC are needed to mitigate the rate of late-stage diagnoses and, likewise, PC mortality. Suitable diagnostic biomarkers require that 1) the molecule is uniquely expressed in the disease setting, 2) the marker is expressed early in the disease establishment, 3) the marker is stably detectable, and 4) sample procurement from patients is non-invasive and readily collectible, ideally from biofluids. Since PC cases in TCGA are comprised of early-stage tumors, exploration of transcriptomic data could identify early detectable biomarkers or prognostically relevant transcripts. Aside from clinical staging and radiographic-based methods currently in use, prognostic molecules may better distinguish those patients more likely to benefit from surgical resection and spare those more likely to experience a rapid post-surgical decline. Additionally, harnessing the power of next-generational sequencing it is possible to screen large databases for potential novel biomarkers.
while also imparting a degree of disease parameter and patient selectivity to improve the predictive power of results. Based on my preliminary assessment and validation of mucin expression at the gene and isoform levels from PC patients in TCGA, expression of MUC4Δ6 contributed to poor survival outcomes in PC patients. Expression of this MUC4 isoform has two implications, 1) the expression of MUC4Δ6 might identify patients with a more aggressive subtype of PC, and 2) the expression of MUC4Δ6 protein might influence the biology or pathology of PC.

Building off my previous results, I have synthesized photoswitchable carboxyfluorescein (FAM)-conjugated DNA oligonucleotides (Fig 4.1) complementary to the MUC4 exon 5-exon 7 junctional sequence using a thiol-bound gold nanoparticle for optical quenching in the unbound state.210 When RNA targets bind the nanoprobes, double-stranded nuclease recognizes perfect hybridization and cleaves the DNA probe, stopping the fluorophore quenching, and permitting its detection and quantification by a plate reader (Fig 4.2). Base mismatches of even a single nucleotide generate bubbles in the double strand and prevent enzymatic cleavage ensuring that fluorophore detection is generated from the target RNA alone (Fig 4.3). The RNA molecule is freed to cycle back and bind another probe in a second order kinetic reaction. The number of probes is uniform across reactions, but the number of matched RNA transcripts is patient specific. The intensity of fluorescence detection is equivalent, but the rate to maximum intensity reveals the concentration of target RNA.

My approach will demonstrate that the detection of circulating alternatively spliced MUC4 RNAs is PC-specific and support their use as biomarkers. The novel
Figure 4.1: Synthesis of RNA-Specific Gold NanoProbes

Probe synthesis begins with synthesis of 13 nm gold nanoseeds capped with citrate (A) and grown to 40 nm (AuNPs) employing the Grabber synthesis technique (B). After confirming integrity and size, AuNPs are PEGylated (C) and concentrated (D). Next, AuNPs are conjugated to PEG linkers (E) and conjugated to bait FAM-DNA oligonucleotides (F) before final washes and concentration (G).
FIGURE 4.1: SYNTHESIS OF RNA-SPECIFIC GOLD NANOPROBES
Figure 4.2: Photoswitchable Fluorescent Probes are Detectable After Enzymatic Cleavage

The detection of target RNA species via my novel gold nanoprobe assay relies on activation of FAM following enzymatic Double-Stranded Nuclease (DSN) cleavage of bait that have perfectly hybridized with target RNA. Imperfect hybridized pairs induce a crinkle that inhibits DSN cleavage.
FIGURE 4.2: PHOTOSWITCHABLE FLUORESCENT PROBES ARE DETECTABLE AFTER ENZYMATIC CLEAVAGE
Figure 4.3: Activation of FAM-Probe Requires Perfect Hybridization of Prey RNA to Bait Sequence

FAM particles are quenched from electron sequestration by the gold nanoparticle. Using an arctic shrimp enzyme DSN, which cleaves perfectly hybridized DNA pairs, target RNA-bait DNA probes are acted on thereby freeing and activating the FAM-DNA from the gold nanocore. Open, non-hybridized segments prevent the enzymatic reaction of DSN, consequently imparting a tremendous degree of specificity to this assay.
FIGURE 4.3: ACTIVATION OF FAM-PROBE REQUIRES PERFECT HYBRIDIZATION OF PREY RNA TO BAIT SEQUENCE
assay presented in this chapter is further strengthened by the fact that it does not require tissue, large sample volumes, complex processing, expensive and tedious sequencing-based techniques, and is scalable for clinical utility.

4.B Results

4.B.1. AuNP Synthesis

Following the protocol detailed in the Methods section of this chapter, 13 nm gold nanoparticles (AuNP) seeds were assessed for size distribution by DLS (Fig 4.4A). Given that DLS measures the particle including its hydroshell, the diameters are slightly larger than labeled with the 13 nm seeds measuring an average of 19.75 nm±5.7 and a polydispersity index (PDI) of 0.094 from more than 2.24*10^7 particles per read. The 40 nm particles (Fig 4.4B) measured an average of 40.38 nm±11.5 and a PDI of 0.096 from more than 1.45*10^7 particles per read. The PDI in both cases indicated a near perfect uniformity of size across analyzed particles.

The concentrations of particles were calculated using a table of nanoparticle physical properties based on particle size and the Beer-Lambert formula (Equation 3) ; where A= the absorbance of light, ε= the molar absorption

EQUATION 3: BEER-LAMBERT EQUATION

\[ A = \varepsilon b C \]
Figure 4.4: Verification of Gold Nanoseeds by DLS

Following the synthesis steps of both 13 nm (A) and 40 nm (B) gold nanoseeds, particles are assessed by dynamic light scattering (DLS) to confirm uniformity of size. Due to the presence of a hydroshell and its optical effect on DLS beams, particle sizes are measured slightly above prescribed sizes. The greater the degree of uniformity across nanoseeds within a single batch, the lower the calculated polydispersity index (PDI) with anything less than 0.1 being acceptable.
FIGURE 4.4: VERIFICATION OF GOLD NANOSEEDS BY DLS

A

13nm AuNP Seeds

Intensity (%)

Z-Average: 17.85nm
Pdl: 0.094
Std Dev: 5.686

Diameter (nm)

peak: 19.75
Read1
Read2
Read3
Average

B

40nm AuNP

Intensity (%)

Z-Average: 40.38nm
Pdl: 0.096
Std Dev: 11.50

Diameter (nm)

peak: 44.67
Read1
Read2
Read3
Average
coefficient, \( b \) = the path length of light, and \( C \) = the concentration of particles. The 13 nm seeds had a max absorbance of 3.1 at 519 nm giving a concentration around 3-4 nM while the 40 nm particles had a max absorbance of 0.55 at 528 nm and a concentration around 8 nM (determination of concentration described in Methods and Materials). The concentrations of all subsequent particle reactions were calculated from the starting concentrations ascertained from the 13 nm and 40 nm UV spectroscopy facilitated calculations.

The AuNPs proceeded to PEGylation (AuNP-Peg) and linker (AuNP-Linker) reactions before the final reaction to conjugate them to the DNA probes (AuNP-Probe). Each batch of particles was assessed again by DLS and UV spectroscopy. In each subsequent reaction, the particles increased their peak absorbance (Fig 4.5A) and particle diameters (Fig 4.5B) with no significant change in maximum wavelength or frequency rates, respectively. In order to assess the true diameter and structure of the nanoparticles (NPs), a diluted sample was mounted on grided silicon TEM wafer (Ted Pella, Inc) for transmission electron microscopy (Fig 4.5C). Nanoparticle tracking analysis revealed that particles spanned a near range around 44 nm. Particles were resolved on a 2% agarose gel to visualize and confirm successful reaction and increases in native particle size (Fig 4.5D).

4.B.2. AuNP-Probe Conjugation Quantification

Following AuNP-Probe synthesis, probes were degraded from particles with DTT as described in Methods. Using a FAM standard curve (Fig 4.6), the
Figure 4.5: Physical Characteristics of AuNPs and AuNP-Probes

(A) UV-visible spectrum. (B) The DLS spectrum of purified 13 nm nanoseeds; 40 nm AuNPs PEG-coated, Linker-bound and DNA conjugated AuNPs. (C) TEM image of AuNP-Probe conjugated particles demonstrates uniform size and dispersion of AuNPs. (D) Image of gel electrophoresis demonstrating successful DNA-AuNP conjugation reaction. The reduced electrophoretic mobility of the AuNP-Probe conjugation compare to the linker conjugated AuNPs reveals the successful immobilization of DNA molecules on AuNPs surface.
FIGURE 4.5: PHYSICAL CHARACTERISTICS OF AUNPROBES

A. Absorbance vs. Wavelength (nm)

B. % Frequency vs. Hydrodynamic diameter (nm)

C. TEM image showing the nanoparticles at 500nm and 100nm scale.

D. Gel electrophoresis showing bands for Link-AuNP, Probe-AuNP, and Link-AuNP.
concentration of DNA probes was calculated by reading the total fluorescence and cross referencing the dilution corrected concentration using the equation of the standard curve to the final concentration of AuNP (Equation 4).

**EQUATION 4: CALCULATION OF PROBES PER AUNP**

\[
N_{Probes/NP} = \frac{[DNA - Probe]}{[AuNP]}
\]

The final calculations are presented using the table below. These calculated concentrations were used to standardize the concentration of DNA-probes utilized in the assay across targets and NP batches (Table 7).

4.B.3. Detection of MUC4 Wild-type vs MUC4Δ6 Synthetic Target

Synthetic RNA oligonucleotides (Integrated DNA Technologies) of MUC4Δ6 were utilized to test the detection of varying concentrations of targets in this assay. Detection of target transcripts by engineered AuNP-Probes follow second order kinetics (Fig 4.7), thus the concentration of target RNA prey alters the slope of the time-resolved fluorescence detection curve but not the peak intensity given enough time for RNA recycling. The kinetic reaction plot of AuNP-Probe-MUC4Δ6 demonstrated a concentration-dependent slopes in FAM detection with the lower limit of detectable concentration of target at 10 pM. RNA oligonucleotides matching the exon 5-exon 6 and exon 6-exon 7 junctional sequences of MUC4 wild-type (MUC4WT) were also designed (Integrated DNA Technologies) to evaluate the
TABLE 7: AUNP CONCENTRATION CALCULATIONS

<table>
<thead>
<tr>
<th>Target</th>
<th>Mole Fraction PEG-NH₂</th>
<th>[DNA-Probe]</th>
<th>Dilution Corrected [DNA-Probe] (x200)</th>
<th>[AuNP]</th>
<th>N_Probes/NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC4Δ6</td>
<td>0.15</td>
<td>2.46±0.01</td>
<td>293±2</td>
<td>5.1</td>
<td>97±0.4</td>
</tr>
<tr>
<td>KRAS^{G12D}</td>
<td>0.15</td>
<td>1.69±0.02</td>
<td>339±4</td>
<td>5.1</td>
<td>66±0.9</td>
</tr>
<tr>
<td>KRAS^{WT}</td>
<td>0.15</td>
<td>1.91±0.13</td>
<td>381±27</td>
<td>5.1</td>
<td>75±5.3</td>
</tr>
</tbody>
</table>
Figure 4.6: FAM Standard Curve Facilitates Calculation of DNA Probes per AuNP

Using a FAM standard plot to generate a linear equation, the concentration of FAM-DNA probes can be calculated. Since the concentration of AuNPs in solution are known, the total number of FAM-DNA bait probes conjugated to each AuNP can be determined in order to normalize the probe load in each reaction batch to batch.
FIGURE 4.6: FAM STANDARD CURVE FACILITATES CALCULATION OF DNA PROBES PER AUNP

FAM Standard Curve

$R^2 = 0.98$
$y = 289x - 227$
### TABLE 8: SYNTHETIC TARGET RNA SEQUENCES

<table>
<thead>
<tr>
<th>Target RNA</th>
<th>RNA Sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC4Δ6</td>
<td>rUrArGrGrUrGrUrGrCrUrCrUrGrArUrArArArUUrG</td>
</tr>
<tr>
<td>MUC4WT Exon5-6</td>
<td>rGrUrCrUrCrUrArUrCrUrCrUrGrArUrArArUrG</td>
</tr>
<tr>
<td>MUC4WT Exon6-7</td>
<td>rUrArGrGrUrGrUrGrCrUrCrCrGrArGrGrGrUrCrC</td>
</tr>
</tbody>
</table>
Figure 4.7: Limit of Detection of MUC4Δ6 RNA with AuNProbe-MUC4Δ6

The AuNP assay for MUC4Δ6 shows robust detection across a range of concentrations. Complete detection of MUC4Δ6 target RNA is inferred by plateauing of the curve. Low concentrations contribute to a low slope of the FAM curve. The concentrations given in the legend are the concentrations of the input sample, which comprise 10% of the entire reaction volume of 40 µL. Therefore, the lower limit of detection given a perfect reaction environment is 10 pM.
FIGURE 4.7: LIMIT OF DETECTION OF MUC4Δ6 RNA WITH AUNPROBE-MUC4Δ6
specificity of the MUC4Δ6 AuNP probe under competitive binding contributed from synthetic wild-type RNA species (Table 8). When compared to these MUC4WT RNAs flanking the exon 6 sequence, detection of MUC4Δ6 RNA demonstrated a high degree of specificity with no significant detection of nontarget RNA using the MUC4Δ6 probe (Fig 4.8). The use of target sequence matched DNA prey instead of RNA result in no difference in detected probe activation. Detection of FAM activation was negligible when target RNAs are in the presence of wild-type RNA indicating no general interference of binding potential imposed by the wild-type sequences. Although it is more likely that this may have been attributed to lack of non-specific enzymatic cleavage rather than differences in binding properties. Potential variations in the interference effect were evaluated using varying concentrations of combinations of target MUC4Δ6 RNA and nontarget MUC4WT RNAs (Fig 4.9). When target MUC4Δ6 RNA were tested at either 100 nM or 1 nM input (10 nM and 100 pM final, respectively), the slopes and maximum RFU intensities were unaffected by the presence of one or both 100 nM MUC4WT (10 nM final) exon 5-6 junction or exon 6-7 junction RNA fragments. These results suggested that the presence of both RNA species in biological fluids would be unlikely to inhibit or interfere with target variant detection when using this assay.

4.B.4. Detection of KRAS Mutations

In similarly described methods, we designed probes to detect KRAS wild-type and the most commonly detected mutation in constitutively active oncogenic KRASG12D. When these AuNP-Probes were evaluated for specificity and sensitivity (Fig 4.10), the intended RNA target species was able to be detected, however, the
Figure 4.8: AuNProbe-MUC4Δ6 Demonstrates A High Specificity to MUC4Δ6 Target

Specificity of the AuNP-Probe assay was evaluated using synthetic RNA and DNA matching MUC4Δ6 and MUC4WT exon 6 flanking junctions. Positive detection was observed only in reactions with the splice variant with minimal background from negative control reactions.
FIGURE 4.8: AUNPROBE-MUC4Δ6 DEMONSTRATES A HIGH SPECIFICITY TO MUC4Δ6 TARGET
Figure 4.9: MUC4WT RNA Does Not Interfere with MUC4Δ6 Detection

To evaluate if MUC4WT exon 6 flanking junctions could interfere with detection of splice variant MUC4Δ6 within a single reaction well, MUC4Δ6 AuNP-Probe was tested against combinations of single or double exon 6-containing wild-type RNA fragments with two concentrations of splice variant. In all combinations, the addition of one or both wild-type species had no interference with the reaction or a reduction of the detected fluorescence.
FIGURE 4.9: MUC4WT RNA DOES NOT INTERFERE WITH MUC4Δ6 DETECTION
Figure 4.10: Validation of KRAS AuNProbe Specificity

AuNP-Probes were synthesized to target and detect constitutively active oncogenic KRAS$^{G12D}$ (A) and KRAS wild-type (B). Both probes were evaluated against RNA of both variants. While both AuNP-Probes were able to detect their intended targets across a range of concentrations from 1 µM to 1 nM, we observed positive detection of nontarget species.
FIGURE 4.10: VALIDATION OF KRAS AUNPROBE SPECIFICITY

A

G12D Probe

- Assay Positive Control
- Kras G12D 1um
- Kras G12D 10um
- Kras WT 10um
- Kras WT 1um
- Kras G12D 1nm
- Kras WT 1nm
- Blank

B

WT Probe

- Assay Positive Control
- Kras WT 10um
- Kras G12D 10um
- Kras G12D 1um
- Kras WT 1um
- Kras G12D 1nm
- Kras WT 1nm
- Blank

RFU vs. Timepoint (min)
nontarget RNA also contributed to positive fluorescent readings. These results were consistently observed across three separate syntheses of KRAS RNAs.

4.B.5. Sample Fragmentation

The large sizes of RNA were determined to cause steric hinderance to the hybridization of probes to target RNA. Therefore, samples containing RNA were fragmented using a salt-based reagent (described in Methods and Materials). To evaluate the optimal fragmentation time, cell line RNA samples were subjected to a series of increasing fragmentations intervals and assessed for RNA quality (Fig 4.11). At a reaction time of 15 min, RNA sizes were around 125-175 bp and determined to be ideal for the application of this AuNP assay. No further significant fragmentation was observed with additional reaction time, so a standard 20 min reaction was elected to tighten the range of fragment sizes.

4.B.6. Detection in Murine Serum

Following the validation of my AuNP-probe assay with synthetic RNA oligonucleotides, its utility with biofluid samples was evaluated. Serum was collected from orthotopic mice implanted with CD18/HPAFII pancreatic cancer cells. All samples were observed to produce positive reads for MUC4Δ6 RNA transcripts except for one animal which produced signal not different from blank and healthy human serum (Fig 4.12). Despite the low concentration of RNA in circulation and the minimal sample used for this assay, a positive detection of MUC4Δ6, which is natively expressed in CD18/HPAFII cells, supported the expansion of this assay to include human PDAC patient samples.
Figure 4.11: COLO457 RNA Fragmentation Validation

RNA from the COLO357 PDAC cell line were isolated and subjected to salt-based fragmentation for 15 min, 2 hr, and overnight to evaluate the optimal fragmentation reaction time.
FIGURE 4.11: COLO357 RNA FRAGMENTATION VALIDATION
Figure 4.12: Detection of MUC4Δ6 in Orthotopic Plasma

(A) Time resolved kinetic reads of fragmented orthotopic plasma demonstrated positive detection of samples, although, the peak amplitudes were exceedingly low compared to positive control reactions. (B) Blanked reads at maximum intensity (occurred at 4 hr and 58 min), samples demonstrated minimal positive signal. (C) Calculation of MUC4Δ6 RNA concentrations in samples revealed detectable traces as low as 0.971 nM in the sample with the lowest signal.
FIGURE 4.12: DETECTION OF MUC4Δ6 IN ORTHOTOPIC PLASMA

**A**

![Graph showing RFU over time for different samples.](image)

**B**

Maximum at 4hr 58min

<table>
<thead>
<tr>
<th>Mean</th>
<th>965</th>
<th>7</th>
<th>53.5</th>
<th>43.5</th>
<th>236</th>
<th>135.5</th>
<th>91</th>
</tr>
</thead>
</table>

**C**

Standard Formula: \( \frac{\text{(RFU} + 227)}{289} = \chi \text{ nM} \)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Max Blanked Intensity</th>
<th>Concentration (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTT</td>
<td>965</td>
<td>4.125</td>
</tr>
<tr>
<td>Healthy Human</td>
<td>7</td>
<td>0.810</td>
</tr>
<tr>
<td>S164</td>
<td>53.5</td>
<td>0.971</td>
</tr>
<tr>
<td>S532</td>
<td>43.5</td>
<td>0.936</td>
</tr>
<tr>
<td>S595</td>
<td>236</td>
<td>1.602</td>
</tr>
<tr>
<td>S595CF</td>
<td>135.5</td>
<td>1.254</td>
</tr>
<tr>
<td>S601</td>
<td>91</td>
<td>1.100</td>
</tr>
</tbody>
</table>
4.B.7. Serum Preparation for Assay Detection

The addition of the fragmentation step was based on the assumption that longer cell-free mRNA species might be inhibited from rapid hybridization due to steric hindrance induced by large secondary structures. The low slope observed from the orthotopic murine serum may have occurred due to few reasons, including low concentration of MUC4Δ6 target RNA, lower reactivity of DSN, or lower than calculated probes per AuNP. The most plausible explanation is that the concentration of MUC4Δ6 RNA was artificially low. Fully exposed cell-free RNA in circulation is likely degraded quickly and is mostly found inside exosomal bodies. Therefore, a new procedural step was necessary to free and stabilize these RNA species. Adapting a biofluid processing protocol from Wang et al. 211, the same orthotopic serum samples along with 3 control samples from saline injected mice were preliminarily prepared using a proteinase K reaction to disrupt exosome membranes and release the sequestered RNA before pelleting membrane and protein components. This prepared sample was used in the AuNProbe assay as isolated with an aliquot proceeding to fragmentation. Both sample preparations were used for the detection of MUC4Δ6 RNA and compared to unprocessed serum (Fig 4.13). This preparation failed to adequately detect the presence of target RNA. Since the proteinase K preparation solution is mixed in a tris-boric acid buffer containing ethylenediaminetetraacetic acid (EDTA), I hypothesized that the magnesium cofactor for DSN was being quenched from the mix and inhibiting DSN reactivity. The samples were prepared again, this time using EDTA-free buffer. As
Figure 4.13: Modification of Orthotopic Plasma Preparation Using Proteinase K Pretreatment

To explore if the low signal-to-noise ratio of the AuNPProbe assay was due to sequestered RNA within circulating exosomes, orthotopic plasma samples were subjected to proteinase K buffer pretreatment alone (Middle), pretreated with proteinase K buffer then fragmented (Bottom), and compared to fragmentation alone (Top). Only fragmentation treatment alone demonstrated a minor positive detection while any proteinase K pretreatment resulted in inverse signal detection.
FIGURE 4.13: MODIFICATION OF ORTHOTOPIC PLASMA PREPARATION USING PROTEINASE K PRETREATMENT
an experimental control, aliquots of orthotopic serum were prepared the initial way and cleaned using a DNA/RNA cleanup protocol.

Following proteinase K preparation with the modified EDTA-free buffer, a positive and amplified signal was detected in the orthotopic plasma samples (Fig 4.14). Concentrations from the previous proteinase K buffer preparations were negative, and actually demonstrated inverse signal curves.

4.C Discussion

The specific and sensitive detection of disease biomarkers is a critical checkpoint that prevents many diagnostic or prognostic assays from achieving validation and progressing to clinical utility. In addition, these clinical assays can typically accompany a high physical cost for patients paid in the contribution of tissue samples collected through costly, invasive, and painful procedures. The innovation of liquid biopsies promises to greatly reduce this burden as many modern platforms can use a variety of biofluids, from urine, to blood or blood fragments, to saliva. Still, early design and standardization approaches in assay development require keen attention to sample processing, assay shelf-life, uniform and replicable results, and minimalization of noise and false signal with increased true detection. However, the fine tuning of true signal to noise becomes an intricate dance between the assay sensitivity and the specificity of the intended detection. Here in this dissertation chapter, I have detailed the conceptualization of a novel activatable gold nanoparticle assay targeting a pernicious RNA splice variant of MUC4: MUC4Δ6, demonstrated earlier in this work to be indicative of an adverse prognostic outcome in patients with pancreatic ductal adenocarcinoma. Our
**Figure 4.14: EDTA-Free Proteinase K Treatment of Orthotopic Plasma**

(A) When orthotopic plasma was pre-processed with proteinase K buffer not containing EDTA, assay reads amplified and demonstrated appropriate plateaus. (B) Notably, the positive control demonstrated a higher maximum concentration owing to the correction in prior EDTA-mediated chelation of magnesium. (C) The calculated concentrations demonstrated about a 2-fold increase in detected concentrations compared to the previous preparation buffer containing EDTA.
FIGURE 4.14: EDTA-FREE PROTEINASE K TREATMENT OF ORTHOTOPIC PLASMA

A

Timepoint (min)

RFU

S601  S532

S595  S164

S595_Cystic Fluid  HealthyHuman

--- Blank

--- Positive Control

B

Maximum at 1hr 29min

Mean

2290  1027.5  1891.5  -60.5  685.5  664.5  297.5

Blanked

2000

1000

0

Positive  S532  S595  S595CF  S601  164

Healthy Human

C

Standard Formula \( \frac{(RFU + 227)}{289} = \chi \text{nM} \)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Max Blanked Intensity</th>
<th>Concentration (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>1610.5</td>
<td>6.358</td>
</tr>
<tr>
<td>S532</td>
<td>207</td>
<td>1.502</td>
</tr>
<tr>
<td>S595</td>
<td>628</td>
<td>2.958</td>
</tr>
<tr>
<td>S601</td>
<td>399</td>
<td>2.166</td>
</tr>
<tr>
<td>S595CF</td>
<td>19</td>
<td>0.851</td>
</tr>
<tr>
<td>164</td>
<td>228</td>
<td>1.574</td>
</tr>
<tr>
<td>Healthy Human</td>
<td>82.5</td>
<td>1.071</td>
</tr>
</tbody>
</table>
innovative assay exhibited a high degree of specific detection which was unaltered by the presence of wild-type MUC4 RNA. Additionally, my assay presented a very strong sensitivity, detecting sample concentrations as low as 10 pM of target MUC4Δ6 RNA. Through the validation of this assay using orthotopic and human MUC4 transgenic animal plasma, it was revealed that additional procedures in sample preparations were compulsory. My work demonstrated that large RNA molecules possess intrinsic spatial properties which inhibit a robust hybridization with the lure DNA probes conjugated to my gold nanocores. To address this issue, a salt-based fragmentation process was conceived and implemented as a core component of the pre-assay preparation. This process consistently reduced RNA species to a near uniform size of around 200 basepair, in effect robustly increasing their binding potential to the nanoprobes. My work also discovered that the concentration of cell-free RNA in systemic circulation is high enough for adequate detection using my assay. However, low detected signal in early experiments also revealed that much of this RNA is locked inside extracellular bodies. Therefore, I also devised a ‘first-step’ process of disrupting the membranes of these exosomal bodies and neutralizing the membrane-integrated proteome using a custom proteinase K buffer. In all, I was able to successful demonstrate that detection of the MUC4Δ6 splice variant RNA is possible and was specific to PDAC-associated plasma samples. Although, the number of healthy or negative control samples employed in this work was limited. Advancing my MUC4Δ6-AuNProbe assay further towards clinical use would require a robust and varied validation in human samples.
I also attempted to detect RNA of the most frequently expressed KRAS mutant: KRAS\textsuperscript{G12D}. Specific DNA-probes complementary to this variant were designed and tested. However, we were unable to meet the high threshold of specific detection of mutant KRAS as the wild-type sequence also contributed positive signal, albeit lower compared to KRAS\textsuperscript{G12D}. Exhaustive literature investigation in the key enzyme enabling this assay to function, DSN, failed to yield insight as to the probable cause of this false detection. Enzymatic function and structural studies on DSN have been dominated on its interaction on DNA-DNA hybrids, which have conclusively identified a 10-basepair binding site for DSN with single strand cleavage between positions 6 and 4. Since our assay functions through RNA-DNA hybridization, the binding properties of DSN are very likely different than those exhibited in DNA-DNA reactions. Some preliminary and unsubstantiated experiments suggested DSN requires a larger perfectly hybridized segment hypothesized to be between 12 and 15 basepair. The lure sequence of my DNA probes is 22 basepair, of which around 15-18 are believed to be accessible to RNA for complementation. Detection of KRAS\textsuperscript{G12D} may benefit changing the length of the DNA probe sequence.

In all, the work presented in this chapter has demonstrated a plausible and promising foundation for a less-invasive diagnostic or prognostic assay detecting alternative splice variants specific to disease states. While much work remains in the optimization and validation of my AuNPProbe assay, the optimistic results of these experiments support the continued investigation of my platform. In addition, I have demonstrated the possibilities of liquid biopsies which will undoubtedly
continue to progress in the future, to the advantage of patients and their care providers and healthcare teams.

4.D Methods and Materials

4.D.1. Synthesis of AuNPs

After ensuring that all glassware was sterilized and washed free of mineral contaminates using Aqua Regia (1:3 nitric acid-hydrochloric acid). All water used throughout the synthesis stages was triple-stage ultrafiltered. Grabar’s synthesis of gold nanoseeds was initiated by bringing 250 mL of water and 2.5 mL 0.1 M gold chloride trihydrate (HAuCl₄, Sigma 52-0918) to reflux in a round bottom boiling flask fitted with a condensation column flowing cold water under constant spinning at ~900 rpm. Upon continuous reflux, 25 mL of 38.8 mM sodium citrate (Sigma S2429) is quickly added to the flask. The solution undergoes a flash reaction turning from pale red to clear then slowly darkens to black/dark blue. The solution is boiled for an additional 10 min then cooled to room temperature. A small aliquot is passed through a 0.45 µm PTFE filter (Acrodisc) and loaded into a 1 cm cuvette for UV spectroscopy (Molecular Devices SpectraMax) and dynamic light scattering (DLS, Malvern Zetasizer Nano ZS) analyses. Seeds that measure ~12-13 nm and have a polydispersity index (PDI) <0.2 are approved to continue through synthesis.

To grow gold nanoparticles (AuNPs), 250 mL of water is added to a glass round bottom boiling flask spinning at ~600 rpm along with 1.5 mL filtered gold nanoseeds. To reduce the gold salt, 1 mL of 40 mM hydroxylamine (Sigma 159417) is added to the flask. Dropwise (approximately 100 mL/hr), 20 mL of 1.25
mM HAuCl₄ solution is added to the flask using an infusion pump (Fisher Scientific 78-01001). After the last HAuCl₄ drop has been dispensed, the solution spins for an additional 1.5 hr before 7 mL of 38.8 mM sodium citrate is added to cap the newly grown ~40 nm AuNPs. These AuNPs are concentrated by ultracentrifugation (Thermo Fisher Sorvall RC 50 Plus) at 10,000 rpm, 4°C for 15 min. The supernatant is carefully removed and AuNPs isolated in a total volume ~2 mL.

The AuNPs are then proceed to PEGylation. In a 5 mL amber glass vial, 870 µL of water is combined with 120 µL of 5 mM SH-PEG-600-NH₂ (Advanced BioChemicals, Lawrenceville, GA), 680 µL of 5 mM SH-PEG-600-COOH (Advanced BioChemicals, Lawrenceville, GA), and the concentrated AuNPs (~6-7 nM). A color change reaction from black/blue to red is observed by adding 330 µL of 300 mM sodium bicarbonate (NaHCO₃, Sigma S5761) solution to the vial before spinning with a magnetic bar for 72 hrs.

Following 72 hr, the reaction solution is added to a 50 kd filter concentrator (Millipore Amicon Ultra-4 50 kd) column pre-flushed with 50 mM NaHCO₃. The filter column is spun at 3000 rpm and 15°C for 5 min (Thermo Fisher Sorvall Legend X1R). A four-step washing procedure follows the initial concentration step with each step spun at 3500 rpm, 15°C for 10 min each. The washing solutions proceed with:

1) 1 mL of 1:1 H₂O-50 mM NaHCO₃
2) 1 mL 4:1 50 mM NaHCO₃-200 proof Ethanol (pure EtOH)
3) Repeat wash 2
4) 1 mL 50 mM NaHCO₃
Following the final wash, the AuNPs are collected in approximately 300 µL and loaded onto a sephadex G-25 purification column (illustra NAP-10) that was pre-flushed with 50 mM NaHCO₃. To run AuNPs through the purification column, 50 mM NaHCO₃ is loaded above the AuNPs in 500 µL boluses. The purified AuNPs are collected in ~500 µL via gravity drip.

Washed and purified AuNPs are conjugated with PEGylated SMCC crosslinker molecules (Thermo Fisher 22103) to prepare for functionalization. In a 1.7 mL microcentrifuge tube, 80 µL water is added to the purified AuNPs, 120 µL of 200 mM phosphate buffer pH 8.5 (Sigma S2429), and 100 µL of 75 mM SM(PEG)₂ linker in dimethylformamide (Sigma 227056) solution under nitrogen gas. The reaction is incubated in a thermoshaker (Eppendorf ThermoMixer C) at 1500 rpm, 4°C for 4 hr. Following the reaction, the AuNP-linker (AuNP-L) solution is loaded into a 50 kd filter concentrator column with 500 µL of 200 mM phosphate buffer pH 8.5 and spun at 3500 rpm, 4°C for 7 min. To remove unreacted AuNPs and linker molecules, a series of washes proceeds as follows:

1) 1 mL of 4:1 200 mM phosphate buffer pH 8.5- pure EtOH
2) Repeat wash 1
3) 1 mL cold H₂O
4) Repeat wash 3

The AuNP-L are collected in ~200 µL and functionalized by conjugation with DNA oligonucleotides complementary to RNA targets.
4.3.2. Functionalization of NPs

Our RNA targets of interest were selected from the bioinformatics analysis from Chapter 1. I elected to target MUC4Δ6, which correlated with increased mortality in TCGA patients. Using a 21-basepair sequence that spanned the exon 5-exon 7 junction site of this transcript, I designed probes with a 5'- 6-carboxyfluorescein (FAM) and 3'-Thiol modifications (Integrated DNA Technologies, Coralville, Iowa). Based on the frequency and increased aggressiveness of PDAC tumors with activating KRAS mutations, I also designed probes to target KRAS wild-type and the most common KRASG12D mutation. The probe sequences designed from bioinformatics in Chapter One and genomic mutation analysis are presented in Table 9.

To prepare the probes for the conjugation reaction, 10 µL of 50 mM tris(2-carboxyethyl)phosphine (TCEP, Sigma 75259) is combined with 10 µL of 400 µM of reconstitute probe to reduce the thiol bond. The reaction is incubated at room temperature in the dark for 1 hr before washing and purification of reduced probe molecules through a NAP-5 sephadex column. Reactive probes are collected in ~200 µL.

In a microcentrifuge tube, 200 µL of 8-10 nM AuNP-L are combined with 80 µL water, 20 µL of 200 mM phosphate buffer pH 7.2, 20 µL of 4 M sodium chloride solution, and the 200 µL purified 200 µM TCEP-treated probe. The final reaction is incubated in a thermoshaker at 800 rpm and 4°C overnight followed by an increase to 1000 rpm for 1 hr to allow probes to bond with the distal end of linker molecules.
# TABLE 9: AUNP PROBE SEQUENCES

<table>
<thead>
<tr>
<th>Probe</th>
<th>Sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC4Δ6</td>
<td>/6-FAM/ CAT TTT ATC AGA GCA ACA CCT A</td>
</tr>
<tr>
<td></td>
<td>/ThioMC3-D/</td>
</tr>
<tr>
<td>KRASWT</td>
<td>/6-FAM/ TTT GGA GCT GGT GGC GTA TTT T</td>
</tr>
<tr>
<td></td>
<td>/ThioMC3-D/</td>
</tr>
<tr>
<td>KRAS\textsuperscript{G12D}</td>
<td>/6-FAM/ TTT GGA GCT GAT GGC GTA TTT T</td>
</tr>
<tr>
<td></td>
<td>/ThioMC3-D/</td>
</tr>
</tbody>
</table>
The nanoprobes (NPs) are pelleted by centrifugation at 10,000 rpm and 4°C for 10 min. The supernatant containing all unreacted probes is decanted and pelleted NPs are resuspended in 500 µL of cold 50 mM NaHCO₃ before filtration through a 50 kd concentrator at 3500 rpm and 4°C for 5 min. Washed and concentrated NPs are washed and purified further by three additional washes of 1 mL 50 mM NaHCO₃. The final purified NPs are collected and resuspended in a total volume of 500 µL 50 mM NaHCO₃ and stored in a dark box at 4°C until use.

4.D.3. Particle Quality Assessment and Quantification

Each batch of NPs contains a similar concentration of AuNPs (3-5 nM), however, the number of probes per AuNP is variable. To begin this calculation, the NP size and concentration are ascertained by comparing the peak wavelength absorption from UV-vis spectroscopy to known values of nanoparticle diameters, concentrations, molar extinction coefficients, and molar concentration. The uniformity of particle size is assessed by DLS analysis. Particle structure is assessed by transmission electron microscopy imaging. The concentration of probes per volume of NPs must be quantified in addition to several particle quality checks. To quantify the concentration of probes, a 20 µL sample of NPs are mixed 1:4 with 1M dithiothreitol (DTT, Thermo Fisher R0861) in 0.2 M phosphate buffer pH 7.8. The NP-DTT sample is agitated at 50°C overnight to completely reduce probes then sonicated for 5 min and centrifuged (10,000 rpm, 10 min). The next day, 20 µL of sample is combined with 180 µL of 50 mM NaHCO₃ and loaded in a 96-well plate. A single timepoint, total fluorescence read is captured for triplicates.
The fluorescence of NPs is compared to a 6-FAM calibration curve to yield the diluted concentration of probes. This concentration is corrected for dilution (by a factor of 200). Using the molar ratio of unknown NP-Probe to previously measured AuNP concentration yields the number of probes per AuNP in solution.


For each reaction, 4 µL of 10x DSN reaction buffer (500 mM Tris-HCl pH 8, 10 mM DTT) is combined with 4 µL 500 mM NaCl, 4 µL 50 mM MgCl₂, 4 U RNase inhibitor (SUPERace-In, Invitrogen AM2694), and a final concentration ~10 nM of DNA-probes on NPs are combined and brought to a total reaction mixture volume of 33 µL with water. The reaction mix is added to a 384-well black plate (Corning 3571) with 4 µL of sample. The wells are mixed and spun down before 3 U of DSN per µL of reaction (ArcticZymes 70600-202) in a total of 3 µL water is added to each reaction well. The wells are then topped with 10 µL silicone oil (Sigma 146153) to suppress evaporation during the reading phase.

4.D.5. Assay Reading

Plates are analyzed in a multimode plate reader (BioTek Synergy Neo2) incubating at 42°C without agitation. Using top optics, fluorescence endpoints are captured for each well at a read height of 4.25 mm and gain of 100. The xenon lamp energy is set to low with excitation wavelength at 484 nm (bandwidth 20) and the emission wavelength detected at 530 nm (bandwidth 25). Readings are taken every 60 seconds over the course of 5 hr for a total of 301 readings.

Because the size of sample RNA imposes a steric hinderance to probe binding, a salt-based nucleic acid fragmentation (Invitrogen AM8740) step was included. For this reaction, 9 µL of sample serum and 9 µL water are mixed with 2 µL of fragmentation reagent and incubated at 70°C for 20 min. After the reaction time, 2 µL of reaction stop reagent are added to each tube and mixed thoroughly. Samples proceed to assay detection immediately after the fragmentation reaction.

4.D.7. Proteinase K Preparation of Serum Samples

Proteinase K solution was prepared by combining 2x tris-boric acid-EDTA buffer with 1.2 M NaCl, 0.4% sodium dodecyl sulfate (SDS), and 25% (w/v) proteinase K (Sigma 1245680500). Serum samples and proteinase K solution were combined in a 1:1 ratio, incubated at 90°C for 5 min, immediately moved to ice for 2 min, and centrifuged at 13,000 x g for 5 min until pellet formed. The RNA-containing supernatant was isolated and used for assay detection or fragmented. After this preparation failed to yield acceptable results, a second preparation was generated exactly as the first except EDTA was eliminated to prevent the chelation of DSN cofactor magnesium from the reaction.

4.D.8. Patient Plasma Samples

Patient plasma samples were isolated and processed from whole blood collected at Nebraska Medicine and stored at -80°C. The collection included
samples from PDAC patients at multiple stages of disease, as well as patients diagnosed with pancreatitis and healthy donors.


Patient samples used in our study were submitted to the UNMC Tissue Bank following proper informed consent procedures to have surplus tissues used for research purposes. Samples were procured by our lab under Aim 1 of IRB Protocol # 186-14-EP. Samples were blinded by central processing upon intake to the tissue bank.
CHAPTER 5:

CONCLUSION
The complex establishment of the pancreas from the foregut epithelium in embryos is a highly regulated, multi-layered, and precisely synchronized process. While not all signaling pathways and molecular crosstalk events are fully described, the complexity and uniqueness of recycled molecules in novel functions regarding pancreatic development demands admiration and respect. From discordant beginnings, as sole foregut epithelium absent a mesenchymal layer, the pancreas is arguably one of the most pivotal evolutionary developments permitting advanced life to arise and flourish. However, loss of homeostasis within the pancreas contributes to diseases that are life altering at best and life-ending at painful worst. Whether loss of $\beta$-cells before or after lost insulin effects in diabetes, or necrosis and fibrosis due to runaway activation of digestive enzymes as seen in pancreatitis, or the myriad alterations and loss or gain of functions of signaling molecules in PDAC, the balance of regulation in the pancreas is poised on a fine point. Any shift in this balance point results in catastrophic consequences and clinical interventions are often unable to reinstate normal functions easily. Diabetes has been one exception in which research and the advent of recombinant insulin were miraculously able to suspend what had once been a death sentence diagnosis. As more work is complete and we uncover more details about the intricate, submicroscopic workings of pancreatic cells, we will undoubtedly expose more questions than conclusions. However, given the changes in cell signals and tissue architecture, one must wonder if some of those answers lie in returning to the beginning, to the remarkable regulation of pancreatic development.
The expression and function of several mucins have been investigated within the context of PDAC; however, the scope of these studies is generally limited to single mucin. A complete exploration of mucin expression in PDAC has not been conducted prior to this work. The establishment of cancer transcriptome databases has increased our ability to assess the potential role of previously unrecognized genes and their variants in specific malignancies. Other studies have reported employing widely used databases to conclude that the expression of MUC4, especially when combined with MUC16 and MUC20, was significantly associated with worse survival outcomes in PC patients. Until present, no study has addressed the need for cellularity-based correction of the RNA-Seq dataset due to the variable and often a low number of malignant cells present in PDAC tissues. This seemingly innocuous oversight obscures the real effects of genes expressed by cancer cells within PDAC tumors. Furthermore, albeit more detrimental, is the indiscriminate use of databases that fail to control for or stratify reported cases adequately. Here I ensured that PDAC samples were correctly vetted before selection. Several groups analyze all PDAC samples in TCGA and inaccurately include samples that were annotated as less than 1% cellularity, neuroendocrine, or primaries of unknown origin, among others. The difference in survival outcome in pancreatic cancer is intrinsic to the type, with PDAC have significantly worse survival. Indiscriminate stratification TCGA samples by mucin expression essentially separates neuroendocrine tumors from PDAC, which the former having substantially better survival.
My work has demonstrated that the proportion of malignant cells present in transcriptome sequenced tumor samples has significant impacts on addressing biostatistical questions. PDAC is well known to contain only a minor malignant compartment with the bulk of the tumor comprised of stromal along with abundant acellular matrix proteins. This concern was addressed by the TCGA Consortium, which used a computational algorithm (ABSOLUTE Scoring) to quantitatively determine the cancer cell population of tumor samples based on gene expression and chromatin methylation states. Based on these scores, they assigned qualitative cellularity labels (high or low cellularity) to each sample. This method, while resourceful, poses a statistical dilemma. While it permits more accurate gene expression analysis from patient to patient, it effectively halves the sample population and reduces statistical power. Because healthy pancreatic tissues are absent of nearly all mucins while PDAC cells express progressive amounts, I elected to use the malignant purity scores of all samples to “normalize” the expression of mucins. In this way, I was able to maintain the full pool of TCGA samples in my statistical analysis and increase the reliability of the results.

In the validation RNA-Seq dataset, however, and like other large cancer databases, the absence of genomic sequencing data prevented the calculation of the ABSOLUTE purity score and, likewise, disallowed malignant cellularity correction or stratification. Consequently, the epithelial samples in this dataset must be comprised of a wide range of malignant cell proportions. Therefore, the expression profiles in this dataset were compared to the TCGA uncorrected mucin expression values; and the epithelial expression was noticeably similar to that
observed in the pooled TCGA uncorrected expression. The observation of epithelial-associated mucin expression in the validation stromal samples was unexpected and could possibly be explained by the communication of mucin RNA out of the tumor and into the surrounding ECM by extracellular vesicles or exosomes.

When I compared the correction-based clustering of mucins in the epithelial samples of the validation dataset to the pooled uncorrected TCGA mucin clusters, there was some observed discordance despite showing similar log-transformed values. In exploring this disharmony, I observed substantial differences in the mapped sequencing depth between the two datasets, with the TCGA samples having between 60M-150M reads while the validation dataset had around 8M-33M reads. This difference may explain why certain mucin expression anticipated to be high, like MUC4, was observed lower than some higher observed mucins, like MUC16, which were expected to be low.

Using the mucin expression profile, I attempted the PDAC subtyping to better understand different routes of disease progression with the chief objective being better therapeutic approaches in the clinic and ultimately improved the survival of patients. Several groups have published works that have significantly advanced the biological understanding of PDAC subtypes. The most accepted subtyping strategies involve complicated and expensive sequencing of tumor samples to either identify pathways impacted by genomic mutations, or transcriptomic profiling of the whole tumor sample or microdissected samples. I have expanded on the well-established aberrant and overexpression of
mucins within PDAC tumors. While it has been documented that specific mucins are expressed at crucial points during PDAC progression, my study was the first to attempt to define disease subtyping based on genes correlated with four groups of co-expressed mucins. Pathway analysis of the correlated genes to each of the four mucin clusters suggests that mucin expression might signal or be involved in unique molecular fingerprints of PDAC tumors. I believe that the expression profile of mucin in PC1 promotes the immunological reaction, as indicated by the many T-cell activation pathways. I have also shown that patients with a high expression of these mucins survive longer than patients with low expression. While the other three types do not significantly impact survival outcomes in patients, they may identify subtypes with which specific treatment options could be more prudent. However, I am constrained by the incomplete data regarding lines of therapy given to the patients in the TCGA dataset. Therefore, more studies are necessary to substantiate my hypothesis.

Numerous studies have reported that malignant tumor cells demonstrate a wide array of abnormal alternative splicing events in their expressed genes, some of which may have novel or unregulated functions, prognostic implications, or diagnostic potential with a clinical impact. Therefore, I employed a focused bioinformatics-based approach to investigate the expression of mucin splice variants in PDAC tumor samples from TCGA patients. Assessing all known mucin transcript variants, I concluded that 12 total mucin SVs, four MUC1, two MUC4, and a single SV each of MUC13, MUC15, MUC16, MUC20, MUC21, and MUC22, have significant associations with survival outcome in patients.
Expression of all MUC1 transcripts revealed improved survival times as did the expression of MUC13 and MUC20, while both MUC4 and the MUC15, MUC16, MUC21, and MUC22 SVs demonstrated decreased survival of PDAC patients. The mechanisms by which these transcripts contribute to changes in survival outcomes are not understood and require future study. Expression of MUC1 is typical of many tissues and cell types, including gastrointestinal epithelium, stromal cells, and immunocytes. Thus, the detection of these transcripts may indicate the presence or activation of cells that impede aggressive disease biology. MUC16, the largest described mucin, is not fully characterized, and its contributions to biological and clinical aspects of PDAC are hypothesized but not well documented. MUC4 is expressed in isolated tissues; however, its expression has been well established in PDAC cell lines and patient tumor samples and absent in healthy pancreas. Other groups have demonstrated that expression of MUC13 is associated with a more aggressive PDAC phenotype in cell line models. However, my analysis from TCGA, as well as within my validation samples, contradicts these findings. This disparate observation may be explained by assessing the functional status of MUC13 during tumorigenesis.

Notwithstanding these observations, no studies have evaluated the tumor-specific role and diagnostic potential of mucin SVs in PDAC. My investigation of mucin transcripts in PDAC transcriptomes demonstrated that splicing of exon six from MUC4 presented with significantly decreased patient survival. This exon codes for the n-terminal sequence of the NIDO domain. Despite an incomplete understanding of the NIDO functionality, it has been linked to interactions between
the expressing cell and the surrounding extracellular matrix. 

Interruption of the NIDO domain may permit loose adherence of PDAC tumor cells and increase their mobility. However, I detected the expression of this SV concurrent with MUC4WT, suggesting that the interaction of MUC4∆6 with MUC4WT may result in reduced patient survival. If true, this further suggests that tumor cells expressing this transcript possess an increased metastatic potential and elevate disease aggressiveness.

I validated the expression of MUC4∆6 and MUC13WT in a separate patient tumor cohort, observing that expression of both genes is higher in samples containing higher malignant cells. Further, I found that high expression of MUC4∆6 is an adverse prognostic marker and presented with significantly shortened survival in my validation PDAC patients. In contrast, expression of MUC13WT was discovered to be a favorable prognostic finding and presented with more prolonged survival. My ability to observe statistically significant differences in survival from TCGA comes from its large sample size. It is worth noting that the overwhelming majority of PDAC patients are diagnosed at late-stage disease compared to the early-stage cases represented in TCGA. Although, the availability of patient samples for my validation studies is significantly limited and may have contributed to the lack of statistical significance in survival time when considering MUC13WT expression. The analysis is likely underpowered to detect differences in survival due to this limitation. Nonetheless, I maintain that expression of MUC4∆6 and MUC13WT are unique to PDAC and should be considered when assessing the outcome expectation of patients.
These bioinformatics results and experimental validations supported the further investigation of mucin splice variants and their roles in PDAC pathology. Because MUC4Δ6 was consistently observed to coincide with a worse prognosis and reduced overall survival, it was selected for additional research. Due to the dismal median survival of patients diagnosed with PDAC, the discovery of novel biomarkers might prove exceptionally beneficial clinically. Therefore, a novel gold nanoparticle-based assay was conceived and developed to detect this variant in patient biofluids, a method most likely to have a significant clinical advantage due to the minimally invasive nature of sample collection. The positive results achieved from this assay might prove impactful enough to support its expansion and redesign for in-clinic screening of suspected PDAC patients or to provide a clearer perspective of disease conditions to the medical provider and patient.
CHAPTER 6:

FUTURE DIRECTIONS
Based on the results obtained from the MUC4Δ6 \textit{in vitro} experiments and the limited literature regarding the putative functions of the NIDO domain, future studies of this splice variant will investigate the possibility of MUC4Δ6 contributing to infiltration of specific tissues. A small collection of studies has suggested that the role of the NIDO domain is to facilitate interaction with the surrounding matrix. Loss of exon 6 in MUC4 brings the tandem repeat and remaining portion of the NIDO domain closer in proximity and may cause the large glycosyl branches to obstruct interaction between NIDO and the ECM. Biophysical studies on domains that contribute to MUC4 speculate that the AMOP domain may have an affinity for interacting with Schwann cells of the neuronal or myelin sheath. Therefore, future studies will investigate the potential enhancement that the inhibition of NIDO-ECM interactions and consequential enhanced mobility of its expressing cell has on AMOP-neuronal interactions and its role in PDAC neuronal invasion. Towards accomplishing these future functional studies, approaches to enriching the expression of MUC4Δ6 in cells must be completed aside from those attempted and presented in Chapter 3.

The negative results of KRAS\textsuperscript{G12D} detection with the AuNP assay described in Chapter Three deserve future consideration. Investigating the potential causes of nonspecific KRASWT and KRAS\textsuperscript{G12D} detection, it was uncovered that the enzyme enabling this assay to function, Double-Stranded Nuclease (DSN), has been poorly investigated in terms of its reactivity against RNA-DNA hybrids and previous work has predominately focused on DNA-DNA activity. DSN is extracted from the hepatopancreas of two arctic marine crustaceans, \textit{Paralithodes}
*camtschaticus* (Kamchatka or Red King crabs) and *Pandalus borealis* (Great Northern Prawn). Studies involving the properties and applications of DSN are nonconverging, dividing important aspects of the enzyme between both species. In experiments involving DSN digestion of DNA-DNA pairs, both enzymes have been found to have a binding site requiring 10 perfectly pair bases, where cleavage occurs between basepair 6 and 4. Further, DSN has been shown to preferentially cleave DNA in a DNA-RNA hybrid. However, the length of perfect hybridization is controversial, falling somewhere between 12-15 basepair. DNA-Probes are designed with 20 bases to yield spacing away from the gold core to facilitate RNA-Probe hybridization and DSN interaction. Thus, a single base mismatch in the KRAS assay may still allow DSN to recognize full pairing on either side of the SNP and cleave the probe from its gold quencher. Future studies should investigate the function of DSN across multiple sizes of RNA-DNA hybrids to better consider the role hybrid length has on enzymatic recognition, binding, and cleavage. A more complete understanding of these aspects will enable adaptations of the aforementioned assay for other future experimental approaches.
References


89. Hegele, Cao, Liu, Costain, Charlton-Menys, Rodger, and Durrington. Sequencing of the Reannotated Lmnb2 Gene Reveals Novel Mutations in


160. Kohlgraf, Gawron, Higashi, Meza, Burdick, Kitajima, Kelly, Caffrey, and Hollingsworth. Contribution of the Muc1 Tandem Repeat and Cytoplasmic...


198. Hossain and Wall. Immunological Evaluation of Recent Muc1 Glycopeptide Cancer Vaccines. Vaccines (Basel), 2016. 4(3).


217. Ng, Menheniott, Every, Stent, Judd, Chionh, Dhar, Komen, Giraud, Wang, Mcguckin, and Sutton. The Muc1 Mucin Protects against Helicobacter

Appendix A: Reproduction Permissions

Appendix A-1: Figure 1.1

---

THESIS COPYRIGHT PERMISSION FORM

Title(s) of the Image(s): Terese Winslow LLC owns the copyright to the following image(s):

Title(s) of Illustration(s): Pancreas Anatomy For the National Cancer Institute 2009 [male GI organs with lower right inset of gallbladder, duodenum, and pancreas body]

Description of the Work: Terese Winslow LLC hereby grants permission to reproduce the above image(s) for use in the work specified:

Thesis title: Mucin Expression Signature and Alternative Splicing in Pancreatic Ductal Adenocarcinoma

University: University of Nebraska Medical Center

Digital Object Identifier (DOI) and Journal ISSN, if available:

License Granted: Terese Winslow LLC hereby grants limited, non-exclusive worldwide print and electronic rights only for use in the Work specified. Terese Winslow LLC grants such rights “AS IS” without representation or warranty of any kind and shall have no liability in connection with such license.

Restrictions: Reproduction for use in any other work, derivative works, or by any third party by manual or electronic methods is prohibited. Ownership of original artwork, copyright, and all rights not specifically transferred herein remain the exclusive property of Terese Winslow L.L.C. Additional license(s) are required for ancillary usage(s).

Credit must be placed adjacent to the image(s) in the following format:

© (copyright year) Terese Winslow LLC, U.S. Govt. has certain rights

Permission granted to:
Name: Christopher M Thompson
Mailing address: 7005 Durham Research Center, 985870 Nebraska Medical Center, Omaha, NE 68198-5870
Email address: c.thompson@unmc.edu
Phone number: 402-559-7754

Signature ___________________________ Date 17 March 2021

Signature ___________________________ Date

Terese Winslow LLC, Medical Illustration
714 South Fairfax Street, Alexandria, Virginia 22314
(703) 836-9121
terese@teresewinslow.com
www.teresewinslow.com
Appendix A-2: Figure 1.2

American Cancer Society

April 20, 2021

COPYRIGHT LICENSE AGREEMENT

Christopher Thompson
University of Nebraska Medical Center
c.thompson@unmc.edu

RI: PRR 75388

Dear Mr. Thompson:

In accordance with the following terms and conditions, the American Cancer Society, Inc. (“ACS”) grants your request to reprint the diagram of the pancreas and islet architecture from the ACS publication “What is Pancreatic Cancer?” as set forth in Exhibit A attached hereto and incorporated herein (the “Material”):

1. The following credit line must be prominently placed on the page[s] in which the Material appears:

“Anatomy of the pancreas. © 2017 the American Cancer Society. Used with permission.”

2. ACS grants Christopher Thompson the one-time, nonexclusive, nontransferable license to reproduce the Material in “Mucin Signatures and Splicing in Pancreatic Ductal Adenocarcinoma”. The Material will not be sold.

3. The rights granted herein are for the English language print, including disability accessible, versions and password protected website only and do not apply to revised editions, foreign language editions, or any other form of print or electronic media. Use of the Material for future reproductions, translations, or any derivative/ancillary works without the express written consent of ACS is prohibited. All rights not expressly granted herein are exclusively reserved to ACS.

4. The Material must appear as set forth in Exhibit A. Deletions, alterations, or other modification of ACS content is strictly prohibited. In addition, any other use of this or other ACS material or information without the express written consent of ACS is prohibited.

The American Cancer Society is the nationwide community-based voluntary health organization dedicated to eliminating cancer as a major health problem by preventing cancer, saving lives, and diminishing suffering from cancer, through research, education, advocacy, and service.

CCLXXIII
Subject: RE: Permission to use image
Date: Tuesday, April 13, 2021 at 12:39:52 PM Central Daylight Time
From: Edward Friedlander
To: Thompson, Christopher M
Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma

Author:

Publication: Cancer Cell
Publisher: Elsevier
Date: 14 August 2017

© 2017 The Author. Published by Elsevier Inc.

Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND)

This article is published under the terms of the Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND).

For non-commercial purposes you may copy and distribute the article, use portions or extracts from the article in other works, and text or data mine the article, provided you do not alter or modify the article without permission from Elsevier. You may also create adaptations of the article for your own personal use only, but not distribute these to others. You must give appropriate credit to the original work, together with a link to the formal publication through the relevant DOI, and a link to the Creative Commons user license above. If changes are permitted, you must indicate if any changes are made but not in any way that suggests the licensor endorses you or your use of the work.

Permission is not required for this non-commercial use. For commercial use please continue to request permission via Rightslink.
Appendix A-6: Figure 1.13

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 13, 2021

This Agreement between University of Nebraska Medical Center -- Christopher Thompson ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number: 5047170815279
License date: Apr 13, 2021
Licensed Content Publisher: Springer Nature
Licensed Content Publication: Nature Medicine
Licensed Content Title: Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy
Licensed Content Author: Eric A Collisson et al
Licensed Content Date: Apr 3, 2011
Type of Use: Thesis/Dissertation
Requestor type: academic/university or research institute
Format: electronic
Appendix A-7: Figure 1.14

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 13, 2021

This Agreement between University of Nebraska Medical Center — Christopher Thompson ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number 5047170900746

License date Apr 13, 2021

Licensed Content Publisher Springer Nature

Licensed Content Publication Nature Genetics

Licensed Content Title Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma

Licensed Content Author Richard A Moffitt et al

Licensed Content Date Sep 7, 2015

Type of Use Thesis/Dissertation

Requestor type academic/university or research institute

Format electronic
Appendix A-8: Figure 1.15
# Appendix A-9: Figure 1.16

**SPRINGER NATURE LICENSE TERMS AND CONDITIONS**

Apr 13, 2021

This Agreement between University of Nebraska Medical Center -- Christopher Thompson ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>5047171155442</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 13, 2021</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Springer Nature</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Nature</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Genomic analyses identify molecular subtypes of pancreatic cancer</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Peter Bailey et al</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Feb 24, 2016</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/university or research institute</td>
</tr>
<tr>
<td>Format</td>
<td>electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
</tbody>
</table>

https://st100.copyright.com/AppPrintableLicenseFrame.jsp?publisher=607-7667-4f00-9f94-0e61666c4edf%2020&targetPage=printablelicense
Appendix A-10: Figure 1.17

AMERICAN ASSOCIATION FOR CANCER RESEARCH LICENSE
TERMS AND CONDITIONS

Apr 13, 2021

This Agreement between University of Nebraska Medical Center -- Christopher Thompson ("You") and American Association for Cancer Research ("American Association for Cancer Research") consists of your license details and the terms and conditions provided by American Association for Cancer Research and Copyright Clearance Center.

License Number 5047171307252

License date Apr 13, 2021

Licensed Content Publisher American Association for Cancer Research

Licensed Content Publication Clinical Cancer Research

Licensed Content Title The Proteomic Landscape of Pancreatic Ductal Adenocarcinoma Liver Metastases Identifies Molecular Subtypes and Associations with Clinical Response


Licensed Content Date Mar 1, 2020

Licensed Content 26