

THE CA19-9 ANTIGEN AND STRA GLYCANS DEFINE INDEPENDENT PANCREATIC  
DUCTAL ADENOCARCINOMA SUBPOPULATIONS IMPROVING DIAGNOSTIC  
ACCURACY AND APPROACH TO PROGNOSTIC CLASSIFICATION

By

Daniel Mark Barnett

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Cell and Molecular Biology – Doctor of Philosophy

2020

## ABSTRACT

# THE CA19-9 ANTIGEN AND STRA GLYCANS DEFINE INDEPENDENT PANCREATIC DUCTAL ADENOCARCINOMA SUBPOPULATIONS IMPROVING DIAGNOSTIC ACCURACY AND APPROACH TO PROGNOSTIC CLASSIFICATION

By

Daniel Mark Barnett

Pancreatic cancer is the third deadliest cancer annually in the United States. Although the vast majority of pancreatic cancer belongs to a single type called pancreatic ductal adenocarcinoma (PDAC), tremendous heterogeneity exists within and between PDACs in their biology and clinical behavior, making it difficult to optimize treatment strategies and therapeutics research. The possibility exists that the heterogeneity results from the fact that PDACs actually encompass several distinct subtypes. Recent research has uncovered much evidence for such subtypes, but so far, the research has not produced clear definitions of the subtypes or associated biomarkers that define them. PDACs express a unique set of glycans derived largely from their origins as duct cells with a protective glycocalyx, including the CA19-9 antigen sialyl-Lewis A (sLeA), which serves as the only approved biomarker of pancreatic cancer, and its near relative sTRA. I hypothesized that the neoplastic cells of pancreatic ductal adenocarcinoma can be separated into subpopulations by their specific glycan expression of sTRA and CA19-9 and that these subpopulations have different functional characteristics and risk for disease dissemination. To test this hypothesis, I used several methods involving both primary specimens and model systems. First, I used multimarker immunofluorescence to detect sTRA and CA19-9 and compare their cellular locations, morphologies, and protein co-expression in tumor and matched adjacent uninvolved tissue, lymph nodes, and metastases. Immunofluorescence was detected by automated microscopy and quantified by novel automated software developed specifically for this project. Clear differences were observed between cancer cells that expressed only CA19-9 and those that expressed only sTRA, as well as a third cell subpopulation represented by dual expression. Dual expression represented a

well differentiated epithelial population of cells in well-formed glandular tissue; CA19-9-only expression represented poor to moderately differentiated cell subpopulations of epithelial and flat (mesenchymal) characteristics; and sTRA-only expression represented poor to moderately differentiated cell subpopulations present in “foamy cytoplasm” and flat (mesenchymal) cell features. The co-expression of MUC5AC and beta-catenin was different between the subsets, indicating differences in differentiation. The differences were preserved in cell-line and patient-derived mouse xenografts. I next tested for differences in metastatic propensity. Xenograft tumors expressing sTRA were more strongly correlated with metastasis than those expressing CA19-9, and primary tumors showed differential correlations with lymph-node or liver metastasis depending on glycan expression. Finally, we tested whether blood plasma levels of these glycans correlate with tissue expression and whether elevations occur in distinct subpopulations of patients. The secretion of glycans into cell-culture media, mouse sera, or plasma from human patients generally correlated with glycan expression in the cancer cells, indicating the value of the glycans as serological biomarkers to indicate the tumor type. Certain tissues expressing only CA19-9 did not secrete to blood plasma, particularly in hyperglandular and very high stromal tissue, suggesting a new cause of false negative CA19-9 patients in PDAC detection. CA19-9 and sTRA were elevated in separate subgroups of patients, each with low false-positive rates. As a result, CA19-9 and sTRA together gave better accuracy of PDAC diagnosis than CA19-9 alone (97% specificity, 65% sensitivity vs. 96% and 46%). In summary, these studies support the concept that distinct subtypes of PDAC can be identified by the expression of sTRA or CA19-9. Additionally, sTRA co-expression with CA19-9 also identified a third subpopulation of PDAC with different morphology, likely aggressiveness, and secretion characteristics. Clinical translation is potentially enabled by the detection of these biomarkers in blood plasma, which provides a new approach to improve diagnosis, prognosis and treatment development.

## ACKNOWLEDGEMENTS

There are so many people to acknowledge for the work contained within and the preparation of this dissertation that they will certainly not all be included here. Thanks for the support and guidance of my research mentor Dr. Brian Haab, without whom I would not have completed or submitted this work. Thank you to my Dr. Ying Liu, Katie Partyka, and Luke Wisniewski, long my closest collaborators in the Haab lab. The Haab lab's external collaborators have been tremendously supportive of our work including but not limited to Dr. Randy Brand (UPMC), Dr. Richard Drake (MUSC), Dr. Aatur Singhi (UPMC), Dr. Herbert Zeh, Dr. Peter Allen (formerly MSK, now Duke University), Dr. Lorenzo Sempere (formerly Van Andel Institute, now MSU), Dr. Tony Hollingsworth (UNMC), and Dr. Anirban Maitra (MD Anderson) as well as the Early Detection Research Network through the National Cancer Institute for their assistance and support in obtaining samples and clinical data. A special thanks is owed to Dr. Ying Huang (Fred Hutchinson Cancer Research Center) for the statistical advice and assistance on analyses as well as study blinding. Another special thanks is owed to Dr. David Monsma for his work, assistance and advice on patient-derived and cell-line xenograft mouse models as well as sample collections along with the Vivarium and Transgenic Core, Quantitative Imaging and Confocal Microscopy Core, and Pathology and Biorepository Core at the Van Andel Institute with special thanks to Kristin Feenstra, Carrie Joynt, Lisa Turner, and Dr. Galen Hostetter. I would also like to acknowledge the support of my thesis committee (Dr. John Wang, Dr. Eran Andrechek, Dr. Andy Amalfitano, Dr. Hua Xiao, and formerly Dr. Chengfeng Yang) and program directors (Dr. Sue Conrad, Dr. Brian Schutte, Dr. Justin McCormick), Dr. Kathy Meek, and support staff (Bethany Heinlen, Becky Mansel, Alaina Burkhardt) for many years of unending support. I must finish with many thanks to my fellow graduate students in the DO-PhD, CMB, and Van Andel Institute Graduate School graduate programs, my family, and the daily support of my wife without whom I would not have completed this work!

# TABLE OF CONTENTS

<b>LIST OF TABLES</b> .....	<b>vii</b>
<b>LIST OF FIGURES</b> .....	<b>viii</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1 Introduction.....	2
1.2 Clinical and Molecular Cancer Screening .....	3
1.3 Clinical Subtypes .....	8
1.4 Pancreatic cancer cell subpopulations.....	10
1.5 Neoplastic Cells in PDAC Precursor Lesions.....	10
1.6 Neoplastic Cells in PDAC .....	13
1.7 Tumor Stroma.....	15
1.8 Pancreatic Cancer Subtypes .....	17
1.9 DNA Subtypes.....	18
1.10 RNA Subtypes.....	20
1.11 miRNA Subtypes .....	21
1.12 Integrated Genomic Characterization Subtypes.....	23
1.13 Protein Subtypes .....	24
1.14 Glycan Subtypes of Pancreatic Cancer.....	25
<b>Chapter 2: The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells</b> .....	<b>29</b>
2.1 Abstract .....	30
2.2 Introduction.....	31
2.3 Results .....	34
2.3.1 The sTRA glycan is elevated in PDAC independently from CA19-9 .....	34
2.3.2 The sTRA and CA19-9 glycans identify spatially and morphologically distinct subsets of cancer cells.....	36
2.3.3 Protein expression differs based on glycan type and differentiation .....	37
2.3.4 The expression patterns of sTRA and CA19-9 predict time-to-progression .....	38
2.4 Discussion .....	40
2.5 Materials and Methods .....	44
2.5.1 Tissue samples and tissue microarrays.....	44
2.5.2 Multimarker immunofluorescence and chemical staining.....	44
2.5.3 Image and data processing .....	45
2.5.4 Statistical analysis.....	46
2.5.5 Biomarker panel selection using MSS .....	46
2.5.6 Patient-derived xenograft (PDX) and cell-line xenograft models.....	47
2.6 Acknowledgements.....	49
<b>Chapter 3: The sTRA Plasma Biomarker: Blinded Validation of Improved Accuracy over CA19-9 in Pancreatic Cancer Diagnosis</b> .....	<b>50</b>
3.1 Translational Relevance .....	51
3.2 Abstract .....	52
3.3 Introduction.....	53
3.4 Methods.....	55
3.4.1 Human specimens .....	55
3.4.2 Sandwich immunoassays.....	55

3.4.3 Statistical methods .....	55
3.5 Results .....	56
3.5.1 Detecting the sTRA and CA19-9 glycans .....	56
3.5.2 The sTRA antigen in CA19-9-negative cancer models and primary tumors.....	57
3.5.3 Improved classification performance using the combined markers .....	57
3.5.4 Blinded validation of improved sensitivity and specificity .....	59
3.6 Discussion .....	60
3.7 Acknowledgements.....	63
<b>Chapter 4: sTRA and CA19-9 expression distinguish independent pancreatic tumor cell subpopulations and prognosis .....</b>	<b>64</b>
4.1 Abstract .....	65
4.2 Introduction.....	66
4.3 Materials and Methods .....	68
4.3.1 Multimarker Immunofluorescence (MMIF) .....	68
4.3.2 SignalFinderIF Software Analysis.....	69
4.3.3 Tissue Microarrays.....	69
4.3.4 Statistical Analysis .....	69
4.4 Results .....	70
4.4.1 Validation of sTRA and CA19-9 elevations as independent indicators of PDAC .....	70
4.4.2 Glycan expression in tumors are correlated with distant recurrence .....	71
4.4.3 High dual expression and high solo expression confer differences in survival in clinical tumors.....	72
4.5 Discussion .....	73
<b>Chapter 5: Conclusions and Future directions .....</b>	<b>77</b>
5.1 Summary .....	78
5.1.1 Definition of Cell Subpopulations.....	79
5.1.2 Plasma sTRA and CA19-9 in the Diagnosis of PDAC .....	80
5.1.3 sTRA and CA19-9 in PDAC Dissemination .....	81
5.1.4 Development and Application of a Multimarker Quantitative Pathology System .....	82
5.2 Future Directions .....	83
5.2.1 Glycan secretion and trafficking .....	83
5.2.2 Improved diagnostics by antibody development and additional glycan biomarker discovery.....	84
5.2.3 Potential Therapeutic Applications .....	86
5.3 Concluding Remarks .....	87
<b>APPENDIX .....</b>	<b>89</b>
<b>REFERENCES .....</b>	<b>179</b>

## LIST OF TABLES

Table 1.1 Incidence and survival of selected common cancers .....	90
Table 1.2 Cancer screening test performance for tests in clinical use .....	90
Table 2.1 Results from 10-fold cross validation .....	94
Table 2A.1 Comparison of biomarker values (averaged over two cores) between tumor and adjacent tissues. ....	103
Table 2A.2 Antibody details .....	103
Table 2A.3 3-marker panel threshold and core averaged marker levels for the images in Figure 2A.5 .....	110
Table 2A.4 3-marker panel threshold and core averaged marker levels for the images in Figure 2A.6 .....	111
Table 3.1 Composition of the sample sets.....	112
Table 3A.1 Training set data .....	128
Table 3A.2 Analysis of covariates between the biomarker data and clinical information.....	137
Table 3A.3 Application of the specificity panel and CA19-9 to blinded samples. ....	141
Table 3A.4 Application of the sensitivity panel and CA19-9 to blinded samples. ....	146
Table 4A.1 Demographic data for patients on clinical tissue microarrays .....	158

## LIST OF FIGURES

Figure 1.1 Model pancreatic cancer microenvironment. ....	92
Figure 1.2 Glycan synthesis pathway for lewis antigens and likely binding partners for CA19-9. .....	93
Figure 2.1 The CA19-9 and sialyl-TRA (sTRA) antigens. ....	95
Figure 2.2 Quantifying signals in tissue microarrays. ....	96
Figure 2.3 Cellular morphologies associated with each glycan.....	97
Figure 2.4 Staining and morphologies in xenografts.....	98
Figure 2.5 Protein expression in various cell types.....	99
Figure 2.6 Associations between glycan type and time-to-progression (TTP).....	100
Figure 2A.1 Additional images from the primary tumors.....	104
Figure 2A.2 Additional images from the cell-line xenografts.....	105
Figure 2A.3 Additional images from the PDX xenografts.....	106
Figure 2A.4 E-cadherin and CK19 expression.....	107
Figure 2A.5 Additional images from each tumor group associated with time-to-progression (TTP).....	108
Figure 2A.6 Images from tumors with misclassified TTP.....	109
Figure 3.1 The CA19-9 and sTRA assays.....	113
Figure 3.2 Complementary elevations of CA19-9 and sTRA in model systems.....	114
Figure 3.3 Complementary elevations in primary tumors and plasma.....	115
Figure 3.4 Biomarker panel development.....	116
Figure 3.5 Application to blinded samples.....	117
Figure 3A.1 Correlations between secreted levels and cellular expression.....	126
Figure 3A.2 Individual marker performance in stage I-II and stage III-IV cancers.....	127
Figure 4.1 The CA19-9 and sTRA glycans and their detection on tissue microarrays.....	151

Figure 4.2 Validation of the CA19-9 and sTRA determined glycotypes. ....	152
Figure 4.3 Tumor lymph nodes and metastases show correlation with their origin tumors. ....	153
Figure 4.4 Survival analysis shows weak anticorrelation for the upper and lower ends of dual expression. ....	154
Figure 4A.1 Normal Distribution testing for metastatic data with normalization.....	155
Figure 4A.2 Regional and distant metastasis glycotype correlations .....	156
Figure 4A.3 Survival distribution and association with glycan expression.....	157
Figure A.1 The SignalFinder system. ....	1755
Figure A.2 Automated image analysis of TMA data.....	1766
Figure A.3 Exploring relationships between markers.....	1777
Figure A.4 Composite images. ....	1788

# Chapter 1: Introduction

## 1.1 Introduction

Pancreatic cancer is currently the third most deadly cancer in the United States and mortality is continuing to rise <sup>1</sup>. Tumor resection has shown the best prognosis of all treatments to date, but few patients are caught early enough and with the right conditions for resection to be attempted. To address this, the field has worked diligently to develop better early detection. To date, there have been no new diagnostics developed to successfully detect tumors early, and no test has firmly established superiority over the only approved biomarker, CA19-9. CA19-9 has both too many false positives and false negatives to accurately serve as a sole diagnostic or screening test, but it has found utility tracking response to therapy.

Once pancreatic tumors are confirmed, they are usually metastatic and no treatment has been shown to sufficiently treat tumors to consistently prevent early death. Survival rates at 5 years have improved to 8%, but most tumors are metastatic at diagnosis (81-90%)<sup>1</sup> and develop resistance to treatment quickly. Tumors detected early often fall in the 10-20% of patients with surgically resectable (local and locally-advanced) disease, where average overall survival is 19 months and 5-year survival is 15-25%<sup>2</sup>. In all others, the current best (first-line) treatment regimens are FOLFIRINOX and gemcitabine with nanoparticle albumin-bound (nab)-paclitaxel where overall survival rates in metastatic disease are 11.4-13.8 months and 9.8-12.1 months, respectively, and average duration to treatment failure is 4.3 months and 3.7 months<sup>3,4</sup>. Despite very short average survival time, there are a small subset of patients that live longer and some very long (>10 years), but currently there are no strong clinical or molecular predictors of long-term patient survival. A few weak indicators (up to 40-50% survival >5 years) have been described and include low detectable CA19-9 (<200U/mL), small tumor size (<20mm), and no invasion of lymph node, nerve, or portal vein in post-resection patients<sup>5</sup>.

In the last decade, there has been extensive work completed in the field to characterize pancreatic tumor morphology, genetics, epigenetics, and metabolism for better understanding of

their biology. Tumor characterization also aims to develop strategies for earlier diagnosis, prognosis prediction, and better treatment efficacy, particularly by identifying informative molecular biomarkers. It is now well known that pancreatic tumors often have low cellularity<sup>6</sup>, live in harsh environments<sup>7</sup>, and are very heterogeneous with clonal cell subpopulations expressing various patterns of gene mutations, as well as RNA and protein expression patterns<sup>8</sup>. With high intratumor heterogeneity, it is likely that not all tumor cells have the same potential for involvement in invasive and metastatic disease that will eventually result in progression and death of patients. I hypothesize that stratification and identification of subpopulations of neoplastic cells with invasive and metastatic potential could allow both better biomarkers for identification and prognosis of disease as well as development of more effective and targeted treatments of pancreatic cancer. Here, I present the current state of molecular characterizations of pancreatic tumors with perspective on how those characterizations represent the biology and clinical state of patients with pancreatic tumors and the state of cancer screening relative to other cancers.

## **1.2 Clinical and Molecular Cancer Screening**

Diagnosis of cancer is a complicated and varied process for most cancers. It may include tests of biological specimens, physical examination, diagnostic imaging, as well as invasive and non-invasive procedures as dictated by symptoms, clinical data, and physician judgement. The reliability and costs of tests are widely variable across cancers and even more so for cancer screening. For cancers with long survival from early detection or where costs of screening tests are low, there has been significant clinical adoption of cancer screening tests. Several examples of screening tests are well-recognized including prostate specific antigen (PSA) urine screening for prostate cancer, guaiac tests and colonoscopies for colon cancer, and mammography for breast cancer. Although significant cancer mortality reduction has been realized, cost benefits of these tests have come into question more recently in the United States, particularly by the

United States Preventative Services Task Force (USPSTF), presenting new cost-benefit considerations for the efficacy and costs of screening tests.

The most prevalent and deadly cancers in the United States are summarized in Table 1.1 and the most common screening modalities for these cancers are summarized in Table 1.2. The sex-specific prostate and breast cancers are the most prevalent cancers (131.1 and 105 cases per 100k)<sup>9</sup> and both have widely adopted screening tests. Both prostate and breast cancers have high overall 5-year survival. Lung cancer remains the most prevalent all-sex cancer (50.99 cases per 100k)<sup>9</sup> in the United States and has a survival profile and disease distribution at diagnosis most similar to pancreatic cancer. Screening by radiography is recommended for the highest risk groups, but poor early detection and poor survival are still evident in the population. Colon cancer is the second most prevalent all-sex cancer with 26.1 cases per 100k<sup>9</sup> and has numerous screening modalities available. Pancreatic cancer has a much lower prevalence of 12.79 cases per 100k,<sup>9</sup> but it is now the third most deadly cancer in the United States due to poor early detection and poor survival regardless of clinical stage at diagnosis. As long adopted standards of cancer screening, clinical performance of screening tests for these cancers should serve as performance benchmarks for the development of new biomarkers and screening modalities.

Prostate cancer is the most common cancer in males with 164,690 estimated new cases in 2018 in the United States and the second most frequent cause of cancer deaths in males with 29,430 deaths expected in 2018.<sup>1</sup> The digital rectal exam has been in regular use for more than a century<sup>10</sup> and has varied estimates of specificity and sensitivity from 40 - 90.7% and 28.6 - 81%, respectively.<sup>11,12</sup> The PSA glycoprotein was discovered in 1971 and gained significant use by the early 1990s as a test for prostate cancer, supplanting Prostate Acid Phosphatase (PAP) as a primary biochemical test for prostate cancer.<sup>13</sup> At a cutoff of 4.0mg/mL, PSA specificity and sensitivity are 93.8% and 20.5%.<sup>14</sup> The American Urological Association recommends a shared decision between patient and physician for patient screening by PSA for men 55-69 with PSA

testing every two years for patients opting for screening.<sup>15</sup> Prostate cancer has 100% 5 year survival when found with only local or regional invasion, but 30% when discovered after distant metastasis.<sup>9</sup> PSA has shown clinical utility in detection of prostate cancer and serves as a valuable benchmark for a successful cancer screening test.

Breast cancer is the most common cancer in females with 266,120 estimated cases in 2018 in the United States and the second most frequent cause of cancer deaths in females with 40,260 deaths expected in 2018.<sup>1</sup> Regular examination by palpation is no longer recommended for all patients, but for patients that express interest, physician instruction is recommended by the American College of Obstetricians and Gynecologists (ACOG).<sup>16</sup> Mammography has been recommended since the 1970s,<sup>17</sup> though criticism has been offered that it may be too sensitive and early cancers that would not develop significant malignancy are overtreated as potentially aggressive cancer.<sup>18</sup> Average cost of mammography in the United States is \$266, though there is considerable regional variation.<sup>19</sup> There is little doubt that breast cancer has seen increased treatment success with mammography, though significant questions remain about whether that success is related to treating cancers that would not advance without treatment. Regardless, 5-year survival of localized breast cancer is 98.7%, regionally invasive is 85.3% and distant metastatic breast cancer is 27.0%,<sup>9</sup> indicating the importance of identifying breast cancers early. Mammography has a sensitivity of 90.5-92.5% and specificity of 83.2-97.9%.<sup>20</sup> Mammography has long served as a benchmark for other cancer screening tests.

Colon cancer is second most common all-sex cancer and second most frequent cause of cancer deaths annually in the United States. There were 97,220 estimated new cases diagnosed and 50,630 estimated deaths in the United States in 2018.<sup>1</sup> Colon cancer is frequently highly treatable when found early with 90.4% 5-year survival for localized disease and 71.4% survival for regionally invasive disease.<sup>9</sup> However, the 23% of cases discovered with distant metastasis only have a 5 year survival of 13.5%,<sup>9</sup> indicating the importance of early detection. The USPSTF recommends using stool fecal occult blood tests (sensitivity 7.2%,

specificity 98.8%) for annual screening of blood in stools, although fecal immunochemical tests (FIT) have shown higher accuracy (sensitivity 23.2-68%, specificity 87.6-97%).<sup>21,22</sup> Colonoscopy is recommended every 10 years for most adults starting at age 50 with high sensitivity and specificity.<sup>23</sup> Early detection and removal of cancerous and pre-cancerous growths by colonoscopy have likely played a significant role in reducing colon cancer mortality by over half in the last 40 years.<sup>9</sup> Colonoscopy is a significant benchmark for cancer screening tests. Colonoscopies have compliance of about 60% due to patient and physician compliance, groups with contraindicating risks for performing colonoscopies, financial barriers and lack of access to services.<sup>24</sup> For these patient populations, new tests have been developed and adopted, such as Cologuard and Epi procolon. Cologuard tests stool samples for multiple DNA target genes and has a specificity of 89.8% and sensitivity of 92.3%, though its performance is noted to be similar to less expensive FIT options in some studies.<sup>25,26</sup> It has received clinical adoption in limited use and has served as a benchmark for other new molecular tests, such as Epi procolon. Epi procolon detects hypermethylation of the SEPT9 gene shed into the bloodstream of colon cancer patients. It has also received FDA approval and has a specificity of 81-90% and sensitivity of 70-73%.<sup>22</sup> These newer blood tests show promise, but best practices and screening algorithms are still being determined for these tests in relation to FOBT, FIT, and colonoscopy.

Lung cancer is the most common all-sex cancer in the United States with 234,030 estimated new cases and the most deadly with 154,050 deaths expected in 2018.<sup>1</sup> Lung cancer incidence in smokers is 1259.3-1308.9 per 100,000 compared to 20.3-25.3 per 100,000 in people who have never smoked.<sup>27</sup> Annual low-dose computed tomography (LDCT) has been used as a screening test for the targeted population of current and previous smokers. In this group, LDCT has a sensitivity and specificity of 80-100% and 28-100%.<sup>28</sup> Although the risk of radiation exposure, cost, and compliance have not been found to be cost beneficial for the general population, the opportunity for early detection in the high risk group of smokers has likely

increased survival by early detection.<sup>28</sup> The mortality of one common type of lung cancer in smokers, small cell lung cancer, has decreased by nearly half since its peak in 1988.<sup>9</sup> Annual chest radiography in lung cancer presents a reasonable benchmark for screening of high risk populations in cancer.

New screening tests are being developed for other cancers, though adoption has not been significant for most. The OVA1 tests has been developed and approved by the FDA for ovarian cancer. It has a specificity of 35-40% and sensitivity of 94-99%.<sup>29</sup> Ovarian cancer has an incidence of 11.49 per 100,000 women and is the seventh most common cancer in women.<sup>9</sup> Mortality is high due to late detection. CA125 has also been used to track disease progression and assist in diagnosis but with a specificity of 77% and sensitivity of 47%, it has only moderate clinical adoption for use in diagnosis, but is regularly tested for pelvic mass due to gynecologic society recommendations.<sup>29</sup> The OVA1 test was approved in 2009 and its adoption is still underway, but it presents a new benchmark for ovarian cancer screening.

Pancreatic cancer is the third most common cause of cancer mortality and does not have a widely accepted screening test. Pancreatic cancer is expected to have 55,440 new cases diagnosed and 44,330 deaths in the United States in 2018.<sup>1</sup> At diagnosis, only 10% of patients have localized disease, 29% of patients have regionally invasive and 52% of patients have distant metastasis, indicating a high need for early detection and diagnosis.<sup>9</sup> Further, 5-year survival in localized disease is 34.3%, regionally invasive disease is 11.5% and distant metastatic disease is 2.7%.<sup>9</sup> CA19-9 is the only FDA-approved biomarker for pancreatic cancer and its sensitivity of 79% and specificity of 82%<sup>30</sup> lags the performance of the screening tests adopted for other cancers. Several new tests have been presented in early stages of research and validation, but as yet, none has achieved clinical adoption or FDA approval.

Cancer screening has seen significant improvement over the last three decades, but pancreatic cancer significantly lags in clinical screening performance relative to other cancers of similar absolute cancer mortality. Although there are no absolute standards, sensitivity or specificity

exceeding 90% with high specificity or sensitivity, respectively, has been important for clinical adoption. There is a significant need to develop better biomarkers or screening modalities with these higher predictive values to detect and diagnose pancreatic cancer. To develop better biomarkers with better detection performance, it is necessary to gain better understanding of the biology of pancreatic tumors to identify new molecular targets for screening, diagnosis and prognosis of pancreatic cancer patients.

### **1.3 Clinical Subtypes**

Malignant pancreatic cancers are divided into four primary clinical types: neuroendocrine, mucinous/cystic adenocarcinomas, acinar cell carcinomas, and pancreatic ductal adenocarcinomas.

Neuroendocrine tumors (PNET) derive from the primary neuroendocrine tissues of the pancreas (islets of Langerhans) and are often also called islet cell tumors. They represent 3-5% of all detected pancreatic tumors,<sup>31</sup> though incidence in autopsy studies was more than 1000 times higher, suggesting very few express clinical symptoms.<sup>32</sup> When these tumors exhibit clinical symptoms, they are divided into two groups: functioning and non-functioning, where functioning tumors express specific hormones (i.e. gastrin, insulin, glucagon, vasoactive intestinal peptide (VIP), or somatostatin) whereas non-functioning tumors do not.<sup>33</sup> Symptoms often appear late in disease progression. The majority of functional neuroendocrine tumors are insulinomas, which present clinically with flushing (i.e. bouts of bodily redness and heat) and hypoglycemia, though nonfunctional tumors are equally common.<sup>32</sup>

Mucinous/cystic adenocarcinomas of the pancreas are characterized by cysts secreting mucin proteins and are usually benign. They are separated into two groups: Mucinous cystic neoplasms (MCN) and intrapapillary mucinous (IPMN) neoplasms. Both are often incidental findings from radiologic imaging and benign conditions (78-90%),<sup>34,35</sup> but share similarities with pancreatic ductal adenocarcinoma when they are malignant.<sup>36</sup> Mucinous and malignant cystic tumors represent 6-7% of all detected pancreatic tumors.<sup>31</sup>

Acinar cell carcinomas are extremely rare (0.3 - 2%)<sup>31,37</sup> and derive from the primary exocrine cells of the pancreas. Acinar neoplasms are primarily characterized by non-specific symptoms such as fatigue, jaundice, abdominal pain, nausea and vomiting.<sup>38</sup> They are largely distinguished by large cells and lacking the major driver mutations of pancreatic ductal adenocarcinomas, though SMAD4 mutations have been observed in a subset of patients.<sup>39</sup> The vast majority (80-90%) of pancreatic cancers are pancreatic ductal adenocarcinomas (PDAC).<sup>31</sup> PDAC tumors are thought to arise through a series of transformation stages in ductal cells known as pancreatic intraepithelial neoplasms (PanINs) or acinar cells that have undergone acinar-to-ductal metaplasia (ADM), after cell reprogramming.<sup>40</sup> Clinically, PDAC tumors often arise in patients who have a history of smoking (attributable in 25% of PDAC), heavy alcohol consumption (>3 drinks per day), and/or chronic pancreatitis.<sup>41</sup> PDAC patients also have a high incidence of diabetes (47%), particularly new onset diabetes (27% of all PDAC).<sup>42</sup> All of these environmental disease predispositions are thought to lead to increased inflammatory states in the pancreas. Patients with a family history of breast, ovarian, colon and pancreatic cancer have an increased incidence of pancreatic cancer.<sup>43</sup> Due to the high risk of both environmental and genetic patient risk groups, there is significant interest in developing screening and diagnostic tests capable of accurately detecting pancreatic cancer in these high risk populations. Each risk factor is also suggestive of biology and genetics with potential for diagnostic and treatment targeting.

Pancreatic ductal adenocarcinoma represents 80-90% of pancreatic cancers, but the remaining pancreatic cancers (i.e. neuroendocrine, mucinous tumors, acinar cell carcinomas and other extremely rare neoplasms) still represent valuable biology and dysregulation of normal tissue maintenance in the pancreas. Normal pancreas, chronically or acutely inflamed pancreas, and precursor lesions (i.e. PanINs) also provide valuable information on the processes leading to pancreatic cancer. Each of these pancreatic states helps to understand the roles of cells and interactions that lead to and sustain cancers of the pancreas.

## **1.4 Pancreatic cancer cell subpopulations**

With high intratumor heterogeneity and multiple potential cells of origin, even in cancer there are likely subpopulations of cells that represent more harmful clinical outcomes for patients. Here, the subpopulations of pancreatic cancer cells and subtypes of pancreatic cancers will be explored for their biological roles in cancer and their impact on outcomes of patients, with a primary focus on pancreatic ductal adenocarcinoma.

A large number of subtypes and tumor cell populations have been reported for pancreatic cancer. These will be considered first by cellular subcompartments (neoplastic cells vs stroma) and then by populations of primary cell types (immune, fibroblasts, other non-cancerous cells, neoplastic, and benign exocrine and endocrine cells) and subpopulations within those cell types in pancreatic cancer. Next, these subpopulations will be considered for their roles in defining subtypes as well as other factors leading to subtype determination. These subtypes will then be considered for their associations and potential significance for outcomes and treatment success.

## **1.5 Neoplastic Cells in PDAC Precursor Lesions**

Morphologically, pancreatic ductal adenocarcinomas are widely viewed as comprising two cellular subcompartments: neoplastic cells and stroma (Figure 1.1). Neoplastic cells derive from epithelial pancreatic exocrine cells or their progenitors<sup>40</sup>. The 2010 WHO consensus classification recognized four types of precursor populations: pancreatic intraepithelial neoplasias (PanINs), Intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm(MCN) and intraductal tubulopapillary neoplasm(ITPN)<sup>44</sup>.

Hruban et al. initially proposed a sequential progression from normal ducts through three stages of PanINs with characteristic genetic alterations before progression to adenocarcinoma<sup>45</sup>. They described PanIN IA as characterized by Her2/neu2 (when present) and KRAS point mutations with ductal cell polarization resulting in a cuboidal to columnar transition with basal nuclei.

PanIN IB is characterized by early formation of papillae in the membrane. PanIN IB/II is

characterized by p16 mutations and increasing papillae (number and size). PanIN III is characterized by increasing pseudocolumnar cells in the increasingly neoplastic ductal wall structure with budding and detaching cell structures. They also described accumulation of combinations of p53, DPC4 (SMAD4), and BRCA2 genetic mutations as driver mutations for the progression of cells through pre-cancerous and cancerous stages. Hruban et al. suggested these transitions and genetic mutations could be used to identify potential diagnostic tests and therapy targets. Early PanINs were proposed to rarely advance into adenocarcinoma, while late PanINs have higher risk of development into adenocarcinoma, though multiple stages of PanIN may be simultaneously present. More recently, the Baltimore consensus has suggested eliminating the PanIN 2 state along with moderate grade designations, downgrading these due to their low likelihood of advancement to carcinoma-in-situ.<sup>46</sup> PanINs may also be generated by nearby PDAC lesions and retain characteristics of their precancerous cells of origin.<sup>47</sup>

In a mouse model of pancreatic cancer, the ER stress response protein, anterior gradient 2 (AGR2), was shown to separate cells of acinar origin and ductal (tubular) origin where acinar cells lead to PanINs, while AGR2-expressing PDAC cells resulted from ductal origin cells that transitioned to PDAC without PanIN stages.<sup>47</sup> In the same study, bystander PanIN lesions were induced by PDAC lesions where cell subpopulations retained their cell of origin AGR2 expression. Similar expression was also demonstrated in clinical samples where AGR2 was shown in cell subpopulations adjacent to tubular and early PanIN lesions.<sup>48</sup> Dumartin et al. also showed ER-stress-induced pancreatic stellate cells induced ER stress, AGR2 expression, and inflammatory marker (IL-6) expression in PDAC cell lines by paracrine signaling. Prior studies have shown that AGR2 is TGF- $\beta$  responsive and suppressed by SMAD4 expression.<sup>49</sup> These authors also showed that MUC1 is exclusively co-expressed with AGR2 and dependent on AGR2 for expression as such suggested that AGR2 and MUC1 are indicators of TGF- $\beta$  transition from tumor suppressor to tumor promoter. These findings suggest that MUC1 may be a potential secreted biomarker and AGR2 as a valuable tissue biomarker of a unique

subpopulation of pancreatic cancer cells of acinar origin with SMAD4 loss in a pro-inflammatory ER-stress state.

Much like PanINs, mucinous cystic neoplasms (MCNs) and intrapapillary mucinous neoplasms (IPMNs) are thought to serve as precursor lesions with sequential genetic mutations leading to PDAC. Both MCNs and IPMNs are heavy mucin-secreting neoplasms, though both are benign until they progress to a PDAC phenotype.

IPMNs have a distinct phenotype marked by mucinous secretion and extensive papillary formation that can often be identified early by diagnostic imaging.<sup>50</sup> IPMNs are divided in clinicopathologic subtypes of intestinal, gastric, pancreatobiliary and oncocytic due to their morphology, protein expression, and behavior. They are initiated by genetic “driver” mutations distinctly different from PDACs, RNF43 and GNAS, though not all clinicopathologic types have the same driver mutation profiles.<sup>51,52</sup> For example, GNAS is present in all intestinal, half of gastric, and 71% of pancreatobiliary, but no oncocytic-type IPMNs.<sup>51</sup> Patients affected by IPMNs with malignancy have much longer survival than other pancreatic cancers, though there is no significant difference in pathological state from other patients’ tumors, suggesting better detection of early lesions.<sup>53</sup> MCNs are cysts that are also mucin-secreting. When sufficiently large, they can also be identified on radiologic imaging, though diagnosis typically requires further diagnostic testing.<sup>54</sup>

Like IPMNs, up to half of all patients with MCNs often have RNF43 mutations.<sup>52</sup> Malignant MCNs also often show KRAS mutations and have genetic (KRAS and RNF43 mutations), morphologic (stromal pattern), and patient demographic (increased risk in females) similarities to mucinous ovarian tumors.<sup>55</sup> It has been hypothesized that these ovarian-type MCNs are germline tumors present from birth and derive from germ cells deposited during development as they traffick to their normal location.<sup>55</sup> If this hypothesis is correct, subsequent studies may elucidate biomarkers and pathways to identify and target this unique cell subpopulation. In other work, aggressive subpopulations of ovarian-type MCNs have been shown to express

carcinoembryonic antigen (CEA), HER2, maspin and neuroendocrine proteins (progesterone receptor, synaptophysin, CD56, and neuron-specific enolase), which together or separately form functional subpopulations.<sup>56</sup>

Like IPMNs, intraductal tubulopapillary neoplasms (ITPNs) are intraductal, but unlike IPMNs and MCNs, they are not mucin-secreting. Histopathology of ITPNs is well-characterized, but as a rare precursor lesion and eventual carcinoma (1% of exocrine tumors), the prognosis and progression of ITPNs are not well understood. ITPNs often present as a solid duct-obstructing mass and share many characteristics with non-ductal pancreatic masses (acinar cell carcinoma and neuroendocrine tumors). Biomarker expression may include cytokeratins, CEA, neuroendocrine proteins (synaptophysin, chromogranin A, somatostatin), exocrine enzymes (e.g. alpha 1 antichymotrypsin), and glycans (e.g. CA19-9/sialyl Lewis A).<sup>57</sup> Though by definition lacking mucins,<sup>44</sup> ITPNs have been shown to be lacking expression in MUC2 and MUC5AC, but frequently express MUC1 and MUC6.<sup>58</sup>

Each of these precursors lesions has a unique genetic expression profile and baseline phenotype from which it begins development into PDAC. Although ultimately converging on a similar end phenotype, the precursor lesions likely provide an imprint reflected in the future adenocarcinoma. Early in development there may even be type switching or early fate-determining steps from a precursor to these precursor lesions.<sup>59</sup> Defining and understanding the subpopulations that form a tumor and the origins of those subpopulations may help to develop better diagnostic and prognostic assays as well as develop more effective treatments.

## **1.6 Neoplastic Cells in PDAC**

Neoplastic cells in PDAC have been shown to be comprised of diverse clonal cell populations spread across wide areas.<sup>8</sup> Pancreatic cancer is highly stromal, which leads to wide spatial separations between clonal cell populations (clones). The fragmentation of clones shows some similarity to fragmentation studied in ecology and evolution. In similarly approaching the pancreas with dysplasia, it has been shown that subclones evolve to form unique

subpopulations but that those subpopulations carry the mutations and genetic marks of parental lines.<sup>8,60</sup> These genetic marks as well as their expressed phenotypes are both important for understanding the biology of tumors and identification of high risk subpopulations.

The presence of several biomarkers has been shown to be indicators of functional subpopulations of neoplastic cells. Dpc4/Smad4 immunostaining was present in 70% of patient tumors with widespread metastasis at death, but not the 30% of patients with localized disease alone.<sup>61</sup> Overexpression and loss of expression of SMAD4 are both correlated with SMAD4 genetic dysregulation<sup>62</sup> and have a significant effect on TGF $\beta$  signaling, a key pathway in pancreatic cancer.<sup>63</sup> Moreover, SMAD4 has a dose response relationship for radiosensitivity with overexpression increasing radiosensitivity and loss making tumors resistant.<sup>64</sup>  $\beta$ -catenin has shown utility, first with upregulation in stressed membranes and then with a nuclear shift as the cell progresses to a more aggressive disease state due to loss of adhesion from separation of the E-cadherin/  $\beta$ -catenin complex.<sup>65</sup>

Cancer stem cells are another subpopulation of interest for pancreatic cancer. If a true cancer stem cell can be identified, both the origins and progression of pancreatic cancer might be better studied and targeted for therapy. The concept of a cancer stem cell in pancreatic cancer is a cell that may be multipotent (multiple potential terminal cell types), self-renewing, malignant and a cell of origin for new neoplastic clones.<sup>66</sup> Potential cancer stem cell subpopulations have been identified with CD133 and they were associated with aggressive proliferation and metastasis.<sup>67</sup> DCLK1 has also been shown to be associated with presumed cancer stem cells with an appearance similar to gastrointestinal tuft cells.<sup>68</sup> CD44 has also been proposed as a marker of pancreatic cancer stem cells as well, and when colocalized with CD133 is found as a marker of centroacinar cells, which is thought to be the progenitor for acinar and ductal cells.<sup>69,70</sup> It is possible that there is more than one subpopulation of pancreatic cancer stem cells that might also address the concerns of lack of expression of all cancer stem cell markers in all

tissues<sup>69</sup> with a convergence of the most successful phenotype in the milieu of pancreatic tumors as a result in any individual patient.

Mucins and glycoproteins may be other useful biomarker to separate subpopulations of neoplastic cells. Several mucin proteins (MUC1, MUC4, MUC5AC, MUC16) and their associated glycans have also been shown to have various functional (gel-formation, structure, receptor binding) and signaling roles (phosphorylation sites and Epidermal Growth Factor domains) both as membrane-bound and secreted proteins.<sup>71</sup> Various glycans, glycan variants, and glycosaminoglycans have been recognized as well, both specifically (i.e. glypican 1, sialyl Lewis A, sialyl Lewis X, and SDC1) and as contributing factors to the functions of glycoproteins (CEA, mucins) in neoplastic cells.<sup>72-75</sup> Combinations of specific mucins, glycoproteins, and their associated glycans may also define more specific subpopulations of pancreatic cancer cells with biological and clinical significance.

These functional proteins and glycans are not expressed in all cells and are often represented in multiple clonal populations within pancreatic tumors. In that state, they may be used as biomarkers to represent a neoplastic state or even a prognostic state based on the outcomes associated with their expression. Further work is needed to associate specific identified subpopulations with their metastatic and outcomes potential for pancreatic tumors.

## **1.7 Tumor Stroma**

In PDAC, stroma often comprises up to 95% of the tumor.<sup>76</sup> PDAC tumors are typically hypovascular and hypoxic. Likely due to their ductal origin and organization, PDAC cells are often well distributed within a tumor and arranged in small pockets or duct-like structures.<sup>77</sup>

Stroma forms the interconnections between all of these distributed centers of neoplastic cells and serves a variety of roles in antagonizing and supporting the neoplastic cells as the regulator of the tumor microenvironment. The biology of tumor stroma and its relative composition have a significant role in determining the phenotype of a tumor. Biomarkers may be used to identify the

state and composition of tumor stroma and provide information on the state of neoplastic cell subpopulations.

Stroma is primarily comprised of fibroblasts (activated and quiescent), immune cells (T cells, B cells, NK cells, macrophages, neutrophils), other non-cancerous cells (endothelial cells and nerve cells in neurovascular bundles), and high deposition of fibrotic extracellular matrix.<sup>78</sup>

PDAC stroma is often organized with fibroblasts and connective tissues, mainly composed of collagen and  $\alpha$  smooth muscle actin depositions, to form high tension in tumors adjacent to cancer cells.<sup>79,80</sup> Fibroblasts are thought to have two states in PDAC tumors, activated and quiescent. Cancer-activated fibroblasts (CAFs) have been proposed to assist in tumor signaling and metabolism of byproducts from tumors.<sup>81</sup> Fibroblasts may also supply critical structure and nutrient supply for the cancer cells within the highly hypoxic environment of most pancreatic tumors.<sup>81</sup> Quiescent fibroblasts serve a more traditional role of maintaining connective tissues within the tumor and likely do not immediately impact the viability and proliferation of cancer cells, though they do store vitamin A and produce matrix metalloproteinases (MMPs).<sup>82</sup>

Fibroblasts serve as the vital core of activity in pancreatic cancer stroma supporting structure, invasion, migration, and metabolism.

Immune cells also serve a large role in stroma, though they are largely quiescent in most tumors. Immune cells sparsely populate central areas in PDAC tumors, especially in advanced tumors, where they have been observed at higher density at the perimeters of tumors.<sup>83</sup> Non-macrophage immune cells are often characterized as depleted, inactive, suppressed, or burned out, though depletion of tumor associated macrophages (TAMs) may restore activity of other immune cells.<sup>84</sup> Immune cells may play an early role in regulating tumor growth, but late stage tumors are largely not impacted by immune response. Some have argued that tumor associated macrophages (TAMs) may actually play a role in stimulating tumor growth and there is strong evidence for their role in immune suppression.<sup>83</sup>

As a tumor compartment, stroma is thought to have roles in signaling, metabolism, and diffusion of drugs to cancer cells. There are several notable biomarkers used to define the state of these characteristics of tumor stroma. Cancer activated fibroblasts (CAFs) are identified by  $\alpha$ SMA, fibroblast activation protein (FAP), and vimentin expression, although vimentin is found on most quiescent fibroblasts, as well.<sup>85,86</sup> Various collagens and the sizes of collagen bands have been characterized for stromal state and even their contributions to prognosis.<sup>79</sup> Tenascin C and SPARC have been identified as indicators of tissue tension,<sup>87</sup> which can indicate tumor resistance to drug diffusion. Immune cells have been characterized for numerous markers of activation state, activity, or cell type. Among these are the markers often used to separate immune cell types, but also cancer relevant markers for T cells (CD3, CD4, CD4, PD-1), macrophages (CD204, PD-L1), B cells (CD20), and NK cells.<sup>88</sup> The relative expression and distribution of these markers in pancreatic tumors can be used to indicate the immune activation or suppression. Each of these markers may be used for phenotype and cell subpopulation characterization to distinguish between cells of different function. Many of these cells have functional roles in interactions with pancreatic cancer cells and together they may help elucidate the tumor-stroma interaction that is thought to drive the advancement of pancreatic cancer.

## **1.8 Pancreatic Cancer Subtypes**

Subtypes of pancreatic cancer can be defined by collections of cell subpopulations or by bulk tumor analysis. Subtype definitions are derived from multiple molecular approaches and can be separated by the type of characterization: DNA sequencing, RNA sequencing, methylation, miRNA, proteomic, and glycomic. Some subtypes are developed by combining multiple modalities, though most are defined by a single modality. These markers alone or in conjunction with other characterization may form the basis for subtypes of pancreatic cancers.

## 1.9 DNA Subtypes

DNA subtypes have been used as one approach to pancreatic cancer subtype identification.

Early studies explored the evolution of cancers and resulted in the identification of potential “driver” mutations necessary to advance progression of cancers to more advanced states. Later studies attempted to identify gene mutations and DNA states to separate tumors by genetic alterations. These studies have provided valuable information on the biology of pancreatic tumors, although they have not yet resulted in changes to treatment.

In 2010, Yachida et al.<sup>8</sup> described clonal subtypes in a precursor to The Cancer Genome Atlas studies. The authors took sequential slices of the pancreas then subsampled the slices to examine the genetics and pathology of pancreatic tumors to describe the variation between neoplastic clones. Metastatic clones were also analyzed and compared to the primary tumor samples. The authors determined that pancreatic tumors have high heterogeneity with consistent driver mutations and both varied and increasing numbers of “progressor” mutations. The metastases showed a very different picture with homogeneous clonality within each metastasis with high concordance to a primary tumor clone, but a variation in which clone provided the likely founder clone for the metastasis. Although they did not identify gene lists to determine particular subpopulations, they proposed clonal associations and evolution of pancreatic cancers are drivers of development, metastasis, and recurrence.

Subsequently, TCGA project produced a new set of subtypes and a different analytical approach to pancreatic cancer in 2015. Waddell et al.<sup>89</sup> showed that a genomic approach could be used to assess pancreatic cancers by looking at 100 PDACs for chromosomal rearrangements by copy number variation and whole-genome sequencing. They confirmed the well characterized mutations of pancreatic cancer (TP53, SMAD4, CDKN2A, ARID1A, and ROBO2) to be particularly susceptible to chromosomal rearrangements. The authors also showed that the minority of patients with inactivation or deficiency of DNA repair and maintenance genes (i.e. BRCA1, BRCA2, and PALB2), which have been previously shown to

be present as susceptibility mutations for PDAC<sup>90</sup>, were 80% responsive to platinum-based therapies. Waddell et al. found the chromosomal rearrangements fit into four subtypes: stable, locally rearranged, scattered, and unstable. The authors defined the stable group (20% of patients) for its relatively low structural variations (<50). They defined the locally rearranged group (30% of patients) for clusters of mutations or amplifications. They noted that these contained rare (1-2%), but potentially actionable changes in some patients (e.g. ERBB2, MET, CDK6, PIK3CA, PIK3R3). Roles for many of these genes/proteins have been previously noted by others<sup>91,92</sup>, but this study validated their rates with a larger sample set (n=100). They defined the scattered subtype (36% of patients) as “a moderate range of non-random chromosomal damage and less than 200 structural variation events.”<sup>89</sup> The unstable subtype (14% of patients) was defined by >200 structural variations. They noted that 10 of 14 samples exhibited a high BRCA signature indicating poor DNA maintenance and a likely driver of this group.

Pancreatic cancer subtypes based on DNA characteristics have provided pertinent findings informing the development of further subtypes, but they have yet to be established as subtypes to inform treatment. The earlier studies that examined the clonal evolution of pancreatic cancer provided useful information on the likely progression of pancreatic cancer. The determination by these studies that metastatic clones developed from diverse primary clones, but that within each metastatic locus there was homogeneity indicated multiple metastasizing events likely occurred suggests that multiple cell subpopulations will likely need to be identified and treated to prevent and eliminate metastases. These early studies also showed a consistent set of “driver” mutations leading cells to an eventual state of progression, suggesting that early cancer and precancer may be broadly treatable and identifiable by cells expressing common pathways. Subsequent studies took a broader genomic approach in an attempt to separate tumors into subtypes by DNA stability and mutational state. In further studies, these DNA states have been noted to exist in a set of tumors with much higher than average cellularity. DNA subtypes have

provided one approach to subtyping pancreatic tumors and the results of their work have been used to guide subsequent studies of pancreatic cancer subtyping.

### **1.10 RNA Subtypes**

Several studies have attempted pancreatic cancer subtyping by RNA expression using primary approaches of microdissection and whole-tumor analysis. Microdissection was initially performed by cutting regions of interest to separate neoplastic cell types from non-neoplastic cell types and in later was performed by computational virtual microdissection. Whole-tumor RNA analysis examined molecular pathways to classify tumor states with biological and treatment relevance. Both approaches identified both biology and potential treatment approaches for pancreatic cancer.

In an initial RNA subtyping study, Collison et al.<sup>93</sup> separated PDAC tumors and cell lines by RNA expression signatures in 2011 and named the subtypes based on their apparent functional expressions of Exocrine-like, Classical, and Quasimesenchymal. Gene lists were determined by non-negative matrix factorization (NMF) and found 62 genes that separated the microdissected PDAC tumors and PDAC cell lines into the three groups. Although the authors were reassured of the validity of the exocrine group by gross immunohistochemical staining for digestive enzymes, they had some skepticism of its validity when the group did not appear in any of their PDAC cell line samples. These first three subtypes and their underlying gene lists have served as a basis for molecular subtyping for subsequent studies in pancreatic cancer. The strength of the Exocrine subgrouping was bolstered by Moffitt et al in 2015,<sup>94</sup> who showed an overlap of 17/17 genes in their own NMF analysis of pancreatic tumors and patient-derived xenografts. Moffitt et al were also able to validate the Collisson classical grouping with an overlap of 20/22 genes. The Moffitt model was developed using a virtual microdissection based on separating stroma from cancer cell expression. However, they identified a completely different third grouping called “basal-like” tumors based on an additional 20 genes and indicated a separation of tumors based on stromal activation state of tumors. They found the

quasimesenchymal grouping represented a split between classical and basal-like when analyzed by their gene set. The Moffitt subtypes have found significant adoption in subsequent studies and represent a more complete view of pancreatic tumors than previous studies with consideration of both neoplastic and stromal tumor compartments.

Subsequently, TCGA analysis by Bailey et al.<sup>63</sup> took a broader view of PDAC and examined whole pancreatic tumors of all subtypes, including tumors with associated mucinous dysplasias (IPMN and MCN) as well as adenosquamous tumors to develop RNA expression subtypes. Perhaps not surprisingly, the adenosquamous tumors grouped into a squamous subtype, IPMNs and mucinous tumors grouped into a pancreatic progenitor subtype, and exocrine-like tumors grouped with an aberrantly differentiated endocrine exocrine (ADEX) subtype, though an additional group called immunogenic subtype for its high activity of immune-related genes suggested a group of tumors with potential for immune activation. This study approached tumors globally. The immune group suggests an increase in immune cells in the tumor, but it also suggests room to increase the activation of those immune cells for treatment of the tumors. Subsequent studies have clinically explored activation of CTLA4 and PD1/PDL1<sup>95</sup>, with unfortunately little positive outcome, to date. This immune-depleted subtype is still an area of active study.

RNA-expression-based subtyping has presented new insight into the genes and pathways that separate tumors with suggestions for potential directions for treatment. It has presented new modes of tumor separation that give quantitative methods for classifying degree of neoplastic and stromal involvement. RNA-expression-based subtyping presents a valuable approach for use in research, although the time to produce results clinically is still too slow for early treatment decisions.

### **1.11 miRNA Subtypes**

miRNA is an RNA type expressed by tumors and in use as a diagnostic or prognostic modality in clinical use. Several groups have published miRNAs associated with pancreatic cancer, but

the only miRNA-based assay with high performance in a large cohort is the panel published by Cote et al.<sup>96</sup> The authors used a panel of miRNAs (miR-10b, -155, -106b, 30c, and 212) for diagnostic discrimination. They showed the panel had 95% sensitivity and 100% specificity. They used chronic pancreatitis (+/- bile duct involvement), healthy controls, and PDAC samples. Cote et al. showed the value of using miRNAs for diagnosis, but they did not perform a blinded validation, and only showed a limited set of controls. Further validation may still prove this miRNA panel as a valuable diagnostic assay.

Cote et al also linked biological relationships for miR-10b (PDAC and CAF cells) and miR-155 (CD45+ T cells), though not in matched tumors expressing these miRNAs. They also noted that miR-21 has biological activity in the same cells as miR-10b, but that miR-21 did not perform well as a plasma biomarker for diagnosis of pancreatic cancer despite previous work showing miR-21 association with poor prognosis in pancreatic cancer<sup>97</sup>. They also were not able to perform a direct comparison to CA19-9 in matched samples. Subsequently, they published a larger set of miRNAs with prognostic value in a prospective cohort [ref Int J Cancer], but these new findings are yet to be validated.

In other unvalidated studies groups of miRNAs have been shown to separate into subgroups with differential survival. One study used 19 miRNAs to separate pancreatic cancers into three subgroups with differential survival [ref namkung j gastroent hepat 2016]. In this study, two significant signaling pathways, p53 and COX2, were impacted by a subset of miRNAs. Another study showed a panel of 13 miRNAs with differences in survival outcomes based on a score developed by relative expression of the 13 miRNAs [oncotarget 2016 zhou et al]. Across all three panels, only miR-106b was included in multiple panels. Although there may be significant value to miRNA subtyping of pancreatic tumors, there is still significant need for further validation of miRNA targets and potential diagnostic and prognostic subtypes.

## 1.12 Integrated Genomic Characterization Subtypes

One potential method to increase the power of molecular characterization for subtype development is to integrate multiple molecular characterization modalities. The TCGA group followed up their analyses of DNA and RNA by synthesizing the results of genomics studies and made a strong point of the role of stroma in the variation of subgroups.<sup>76</sup> The authors cited the 2012 Wood and Hruban<sup>77</sup> study for determining “5-20% of neoplastic cellularity” of primary pancreatic tumors, while genomic studies, such as Waddell et al., selected only tumors over 40% cellularity for analysis. The implication is that true subgroups may be underrecognized by the biased population of analyzed tumors. This may also contribute to the stronger alignment of subsequent studies with subtypes using broader sampling and true or virtual microdissection (Moffitt, Bailey). After reclassifying neoplastic cellularity as “purity,” the authors assessed the various described subtypes by purity in a group of 150 tumors consisting of 76 high purity and 74 low purity tumors. They showed that Moffitt subtypes segregated equally between basal and classical types by purity. Collison subtypes showed higher purity in classical, but exocrine and quasimesenchymal showed low purity. Bailey types showed higher purity tumors segregated into squamous and progenitor types, while ADEX and immunogenic were both low purity subtypes. This study then attempted to integrate miRNA analysis based on previous studies, but clustering was a little weak. The subgroups identified only showed a significant separation between the three clustered groups with miR21, which has previously been shown to be associated with prognosis. Cluster group 2 also showed a higher mutation rate in previously identified driver mutations. Further, cross-platform analysis with Similarity Network Fusion (cluster of clusters analysis), showed miR 192 and miR 194 as integrally associated with classical vs basal tumor status.

Whole genome sequencing required for genomic and transcriptomic subtyping is slow clinically. This presents a current weakness in using genomic data for subtyping and clinical decision-

making as data is not usually available until the second- or third-line treatment decision in pancreatic cancer.

### **1.13 Protein Subtypes**

Molecular subtyping of pancreatic cancers has been a goal of pancreatic cancer researchers for at least two decades. In the 2000s, Hruban and Adsay defined molecular characteristics for a set of subtypes of pancreatic cancer based largely on protein expression.<sup>36</sup> Precursor lesions and malignant cancers often contain the same pancreatic cancer driver mutations. Iacobuzio Donahue et al cancer evolution studies showed that cells expressing favorable mutations for the environmental constraints are selected and enriched as precursors lead to cancer and cancer becomes more advanced and aggressive<sup>8,61</sup>. Various protein immunolabeling biomarkers show significant associations with pancreatic cancer subtypes. E-cadherin indicates moderate- to well-differentiated cancers, while loss of E-cadherin indicates poor differentiation. Similar to E-cadherin, loss of  $\beta$ -catenin also indicates loss of cell-cell adhesion and progression to poorer differentiation. Both are associated with poorer prognosis. A series of markers (CD10,  $\alpha$ -1 antitrypsin, vimentin, neuron-specific enolase, and progesterone receptor) are associated with “foam cells, clear cells, cholesterol clefts, and eosinophilic hyaline globules”<sup>36</sup> in solid-pseudopapillary neoplasms, which carry better prognosis than other malignant neoplasms. More, these tumors have a unique loss of cohesion that associates with loss of both E-cadherin and a shift in  $\beta$ -catenin from the membrane to the cytoplasm and nucleus, which further associates with increases in c-myc and cyclin D1. By protein immunolabeling, the biology and progression of pancreatic tumors have become better understood.

These studies explored the use of one to two modalities to define subtypes of pancreatic cancer with biological and clinical relevance. A future goal for studies should be to combine significant biomarkers of each modality in an integrated analysis, much as the recent integrated genomic analysis from the TCGA consortium did with genetic data.

## 1.14 Glycan Subtypes of Pancreatic Cancer

Although there has been significant recent expansion in attempts to develop genetic biomarkers for pancreatic cancer, the only widely-adopted and approved biomarker for pancreatic cancer is a glycan. Glycans are present on the surface of many cell types and many secreted proteins. In normal pancreatic tissues, glycans are expressed on the glycocalyx of pancreatic ducts.

Pancreatic ducts in normal pancreas transport digestive enzymes from the producing acini to the common bile duct for secretion to the intestines. The glycocalyx provides a physical barrier for tissue to prevent damage from digestive enzymes. When dysplasia develops in the pancreas, it is frequently in pancreatic ducts where mucin proteins are widely expressed with glycan motifs. Although several mucin proteins have been characterized as biomarkers, none have been shown to be as frequently or consistently expressed in pancreatic tumors as the most frequent glycan motifs. Most glycans have not been well-characterized for functional significance, but several glycans have been shown to have biological function in non-cancer systems. Due to their frequent expression in pancreatic cancers, glycan biomarkers have strong potential for use in detection and subtyping of pancreatic tumors and cancer cells, and some may have potential biological functions that could be exploited in the study and treatment of pancreatic tumors.

Glycan biomarkers present additional opportunities for pancreatic subpopulation identification and subtype, but functional characterizations and outcomes of specific glycan-expressing subpopulations are lacking in the understanding of pancreatic cancer. The only currently approved biomarker for pancreatic cancer is the CA19-9 antigen, which is the sialyl Lewis A (sLeA) glycan (Figure 2). sLeA is closely related to the blood group antigens including A, B, and H<sup>98</sup> and belongs to a family of glycans called the Lewis glycans. Of clinical and biological relevance, it is also closely related to the E selectin-binding epitope sialyl Lewis X (sLeX), the neutrophil antigen Lewis X (LeX, CD15) and the human embryonal stem cell biomarkers TRA 1-

60 and TRA 1-81<sup>98</sup>. Although expression of these glycans has been characterized in normal tissues, their presence and functional roles in tumors is poorly characterized.

Though they have shown promise for limited application for diagnosis and help inform the larger clinical picture of PDAC tumors, current glycan biomarkers of pancreatic cancer have several limitations for clinical use. As the only approved biomarker, CA19-9 is not very predictive of disease progression with a drop of 20-50% in blood serum resulting in improved prognosis<sup>99</sup>. CA19-9 alone also results in both too many false positives and false negatives for diagnosis or screening due to prevalence in chronic benign conditions and lack of expression in many individuals.

The absence of CA19-9 in some patients has two primary causes: 1. The lack of the enzyme to add the final monosaccharide to produce it, and 2. competition for its precursor by upregulation of another glycan in the synthesis pathway. CA19-9 is expressed in 80-90% of PDAC patients<sup>100</sup>. Of the 10-20% of patients lacking CA19-9, 7-10% lack the fucosyltransferase 3 (FUT3) enzyme required for most sLeA production<sup>101-103</sup>, represented in Figure 1.2A. Another 19.7-22.5% of patients have secretor status where fucosyltransferase 2 (FUT2) consumes the precursor to sLeA<sup>101,103</sup>, significantly reducing sLeA expression. Although not as well characterized, sLeX and the sialylated variant of TRA 1-60 (sTRA) have also been shown to be elevated in blood serum of PDAC patients, both with and without sLeA<sup>104</sup>. sLeX and sTRA represent an alternate precursor and incomplete synthesis of sLeA, respectively. They may also account for some of the patients lacking CA19-9.

With the primary group of patients lacking CA19-9 coming from patients lacking FUT3, detection of sTRA is a prime candidate to reduce false negatives in the detection of PDAC. sLeA is a terminal tetrasaccharide Type I LacNAc consisting of Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3(Fuc $\alpha$ 1-4)GlcNAc (Figure 1.2B) that is displayed at the terminal ends of glycan chains on proteins and potentially glycolipids<sup>105</sup>. sTRA is detected indirectly by applying a sialidase treatment to a sample and detecting with the TRA 1-60 antibody<sup>104</sup>. The TRA 1-60 antigen was initially described as a

terminal structure of keratin sulfate (KS) on podocalyxin in human teratocarcinomas<sup>106,107</sup>.

Glycan array data has since shown TRA 1-60 binds very specifically to a terminal tetrasaccharide with a Type 1 and Type 2 LacNAc with high specificity for the N-Acetyl group on the glycan root glucose and no terminal sialic acid<sup>108</sup>. This presents a structure of Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4GlcNAc and an sTRA implied structure of Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4GlcNAc with terminal structural similarity to sLeA without fucose (Figure 1.2B). Due to the terminal end structure of sTRA containing the sLeA antigen without the fucose added by FUT3, it represents a prime candidate for a biomarker to reduce false negatives of CA19-9. Although CA19-9 has long been characterized to detect sLeA, it likely also binds fucosylated sTRA containing sLeA, which further suggests a potential valuable role for sTRA in PDAC. The Consortium for Functional Glycomics glycan array data shows that all tested CA19-9 antibody clones bind a fucosylated glycan variant of sTRA Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3(Fuc $\alpha$ 1-4)GlcNAc $\beta$ 1-3Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc<sup>109-111</sup> (Figure 1.2B) with similar to higher affinity than terminal sLeA alone. This structure contains the sTRA glycan with two fucose modifications,  $\alpha$ 1-4 on the Type 1 LacNAc and  $\alpha$ 1-3 on the Type 2 LacNAc. Only one antibody shows very low binding to the sTRA glycan itself. This suggests that the primary CA19-9 ligand in tissue could be the fucosylated sTRA glycan rather than the shorter sLeA terminal tetrasaccharide. Also of note, all of the antibodies bind the fucosylated, sialylated Type I-Type I polylectosamine (Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3(Fuc $\alpha$ 1-4)GlcNAc $\beta$ 1-3Gal $\beta$ 1-3(Fuc $\alpha$ 1-3)GlcNAc) with equally high affinity as the sTRA variant. CA19-9 antibody binding to these glycans suggests both sTRA and sialyl polylectosamine (Type 1) may be informative for pancreatic cancer diagnosis, particularly in Lewis-negative patients lacking FUT3 or secretors with high FUT2.

Over the next three chapters, I will examine the importance of cancer cell subpopulations and the use of glycan biomarkers to identify them. I hypothesize that the expression of sTRA and CA19-9 glycans represent two separate subpopulations of pancreatic cancer cells and these subpopulations have different phenotypes for aggressiveness of pancreatic cancer.

To examine this hypothesis, I tested pancreatic ductal adenocarcinoma (PDAC) tumors and matched adjacent uninvolved tissues to quantify differences in CA19-9 and sTRA expression by quantitative immunofluorescence. After demonstrating significant differences between tumor and normal for CA19-9 and sTRA, I examined the spatial distributions of both markers in neoplastic pancreatic cells. I determined sTRA and CA19-9 had significant differences in spatial and morphological distribution both individually and together as dual expression, implying the existence of three cell subpopulations. In these populations, I compared the tissue expression of the glycans to matched serum values and examined secretion to determine the value and constraints of these glycans as biomarkers. CA19-9 and sTRA were then tested as independent and combined biomarkers for pancreatic cancer detection. As validated biomarkers with distinct morphological characteristics, the markers were tested in additional tissues to validate preliminary findings and then in matched tumors and metastases to determine whether there were differences in trafficking and establishment of metastasis. Tissue glycan expression was then tested for prognostic value with progression-free survival in one sample set and overall survival in a second sample set. Together these measures asked whether sTRA and CA19-9 represent independent subpopulations of neoplastic cells and whether those subpopulations express different phenotypes affecting pancreatic cancer progression and outcome.

## Chapter 2: The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells

Daniel Barnett,<sup>1#</sup> Ying Liu,<sup>1#</sup> Katie Partyka,<sup>1</sup> Ying Huang,<sup>2</sup> Huiyuan Tang,<sup>1</sup> Galen Hostetter,<sup>1</sup> Randall E. Brand,<sup>3</sup> Aatur D. Singhi,<sup>3</sup> Richard R. Drake,<sup>4</sup> and Brian B. Haab<sup>1\*</sup>

<sup>1</sup>Van Andel Research Institute, Grand Rapids, MI

<sup>2</sup>Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>3</sup>University of Pittsburgh Medical Center, Pittsburgh, PA

<sup>4</sup>Medical University of South Carolina, Charleston, SC

\*Correspondence:

Brian B. Haab, PhD

Van Andel Research Institute

333 Bostwick NE

Grand Rapids, MI 49503

[brian.haab@vai.org](mailto:brian.haab@vai.org)

#Contributed equally

Published in Scientific Reports 22 June 2017

*Scientific Reports* Vol 7, Article number: 4020 (2017)

## 2.1 Abstract

Molecular markers to detect subtypes of cancer cells could facilitate more effective treatment. We recently identified a carbohydrate antigen, named sTRA, that is as accurate a serological biomarker of pancreatic cancer as the cancer antigen CA19-9. We hypothesized that the cancer cells producing sTRA are a different subpopulation than those producing CA19-9. The sTRA glycan was significantly elevated in tumor tissue relative to adjacent pancreatic tissue in 3 separate tissue microarrays covering 38 patients. The morphologies of the cancer cells varied in association with glycan expression. Cells with dual staining of both markers tended to be in well-to-moderately differentiated glands with nuclear polarization, but exclusive sTRA staining was present in small clusters of cells with poor differentiation and large vacuoles, or in small and ill-defined glands. Patients with higher dual-staining of CA19-9 and sTRA had statistically longer time-to-progression after surgery. Patients with short time-to-progression (<2 years) had either low levels of the dual-stained cells or high levels of single-stained cells, and such patterns differentiated short from long time-to-progression with 90% (27/30) sensitivity and 80% (12/15) specificity. The sTRA and CA19-9 glycans define separate subpopulations of cancer cells and could together have value for classifying subtypes of pancreatic adenocarcinoma.

## 2.2 Introduction

Pancreatic cancers display significant diversity in their rates of growth and dissemination and in their responses to drugs, leading to uncertainty in determining the best treatment for each patient. Methods to subclassify pancreatic cancers to predict behavior clearly would be valuable both for patient care and drug research. At the histomorphological level, several variants of pancreatic ductal adenocarcinoma (PDAC) have clinical implications<sup>36</sup>. For example, medullary adenocarcinomas usually are microsatellite-unstable and have better prognoses<sup>112</sup>; adenosquamous cancers may respond better to platinum-based agents<sup>113</sup>; colloid cancers typically are less aggressive<sup>114</sup>; undifferentiated carcinomas frequently have amplifications of mutant *KRAS* and are more aggressive<sup>115</sup>, and undifferentiated with osteoclast-like giant cells may have better prognoses than conventional PDAC<sup>116</sup>. These variants, however, are the minority; most are conventional PDAC. Further subtypes of PDAC may require molecular biomarkers for identification, as they would be largely indistinguishable by morphology.

Stratification by particular DNA mutations may provide additional guidance in treatment. Mutations in DNA repair genes, such as *BRCA2* or *PALB2*, are often sensitive to poly(ADP-ribose) polymerase (PARP) inhibitors and cisplatin<sup>117,118</sup>, and mutations in the mismatch repair genes confer increased susceptibility to immune checkpoint inhibitors<sup>119</sup>. Subtyping by genome-wide signatures also has shown promise for subtyping tumors. Recent studies identified recurrent classes that were distinguished by genes relating to development, differentiation, and immune infiltration<sup>63,94,120,121</sup>, and a squamous class showed shorter survival<sup>121</sup>. Nevertheless, further research is needed to link practical biomarkers with cancer behavior.

One of the difficulties associated with the molecular profiling of pancreatic cancers is heterogeneity in the tumors. Multiple types of cells may be present at variable levels, all amidst hugely varying backgrounds of extracellular matrix. With methods that use homogenized tissue, one cannot determine which cells produce each marker, or whether certain cells co-express

various markers. A bioinformatics method could sort out the information indirectly<sup>94</sup>, but with limited precision. An additional challenge stems from the possibility that more than one subpopulation of cancer cell could coexist in a tumor. In support of this concept, studies involving isolations of tumor cells with particular stem-cell antigens suggest a minority subtype with heightened tumor-forming capability<sup>122,123</sup>, and analyses of tumor cells in the blood suggest a subpopulation that is able to disseminate prior to clinical manifestations of a primary tumor<sup>124,125</sup>. As indirect evidence, the fact that chemotherapy often reduces tumor volume without eliminating the cancer suggests a subpopulation of cancer cells that is more resistant to treatment than the rest.

To distinguish between individual cells in their expression of one or more markers, a cell-by-cell analysis is required. In the present research we took such an approach using multimarker immunofluorescence (IF)<sup>126</sup>. The method involves probing a single section of formalin-fixed, paraffin-embedded (FFPE) tissue with multiple rounds of multispectral immunofluorescence, each round involving two or more unique antibodies. Multimarker IF gives direct observation of the locations and morphologies of the cells producing each marker, and it is compatible with FFPE tissue, which is easier to obtain than frozen tissue.

We were particularly interested in glycan expression. In previous research we identified a glycan that is a strong serological biomarker of pancreatic cancer<sup>127</sup>. It performed as well as the current best serological biomarker for pancreatic cancer, CA19-9, which also detects a glycan, and it was elevated in about half of the patients with low CA19-9, indicating independent regulation. These facts led us to speculate that the glycan, which we call sTRA, is produced by a different subpopulation of cancer cells than produce the CA19-9 antigen. To test that hypothesis, we sought to immunologically detect sTRA and CA19-9 in tumor tissue and test for differences in location, morphology, and molecular expression of the cancer cells that produce each glycan.

Furthermore, we asked whether particular glycan levels show an association with the rate of progression of pancreatic cancer.

## 2.3 Results

### 2.3.1 The sTRA glycan is elevated in PDAC independently from CA19-9

The CA19-9 antigen is a tetrasaccharide (Fig. 2.1A) that can be detected with high specificity using monoclonal antibodies (mAbs). The TRA-1-60 and TRA-1-81 mAbs<sup>128</sup> detect a tetrasaccharide that, unlike the CA19-9 antigen, is neither fucosylated nor sialylated<sup>108</sup>. To indirectly detect the sialylated version of the TRA antigen (referred to as sTRA), we incubate the labeled TRA mAb to detect and mask the non-sialylated antigens, treat with sialidase, and again incubate the labeled TRA mAb (Fig. 2.1A) to detect the newly-exposed antigens. A view of the structures shows the similarity between the CA19-9 and sTRA antigens, as well as their main difference of a branched fucose on CA19-9 (Fig. 2.1B).

The treatment of tumor tissue with sialidase markedly increased staining by the TRA antibody (Fig. 2.1C), indicating higher levels of the sialylated antigen relative to the non-sialylated. The central question explored here is whether the cancer cells producing sTRA are different in their locations and characteristics than the cancer cells producing the CA19-9 antigen (referred to simply as CA19-9).

To probe this question in primary tissue we chose multimarker immunofluorescence, which allows for the performance of multiple antibody incubations and staining with hematoxylin and eosin (H&E) on a single section from FFPE tissue. We performed three rounds of immunofluorescence, in each round detecting blue fluorescence from a DNA stain, green fluorescence from a Cy3-labeled antibody against a protein, and red fluorescence from a Cy5-labeled protein against a glycan (Fig. 2.1D). We applied the TRA and CA19-9 mAbs in the first and second rounds, respectively, treated the section with sialidase, and then applied the TRA mAb again in the third round.

We applied the method to six separate TMAs, four made from primary tumors and two made from xenograft tumors (Fig. 2.2A). We acquired tiled images across the entire TMA. The field-of-

view of each image was 500 x 400  $\mu\text{m}$ , requiring 6-9 images to cover a core (Fig. 2.2B). Each field-of-view comprised a stack of 35 images collected at various wavelengths, from which we selected the three images corresponding to the fluorescent dyes used here. We quantified the amounts of signal using custom software that employs the SFT signal-finding algorithm<sup>129</sup>, and we then quantified the relationships between the signals from each color. Of particular interest was the possibility that the exclusive expression of a particular marker, i.e. the presence of one marker in the absence of another, could be a marker of phenotype. We therefore designed software to quantify exclusive expression as well as colocalization (Fig. 2.2B).

We first wanted to know which glycans or glycan combinations are elevated in pancreatic tumors relative to adjacent tissue from the pancreas. We examined the signals from CA19-9, sTRA, CA19-9 in the absence of sTRA (referred to as CA19-9-only), sTRA in the absence of CA19-9 (referred to as sTRA-only), and colocalized expression of both CA19-9 and sTRA (referred to as dual-labeling). Both CA19-9 and sTRA were significantly elevated in the tumors, as were the exclusive and dual expression of the markers in most cases (Fig. 2.2D). Using combined data from TMAs 2, 5, and 6, each of the five markers was significantly elevated ( $p < 0.001$  based on Wilcoxon signed rank test, with false discovery rate  $< 0.001$  accounting for the five markers tested) in the tumors relative to paired adjacent tissue (Table 2A.1). Without sialidase pretreatment, detection with the TRA-1-60 mAb did not show significant elevations in the tumors (not shown).

We asked whether the sTRA expression occasionally occurs in locations and tumors that are separate from CA19-9, or whether it is simply an overlapping subset of CA19-9 expression. To begin, we recorded how often a tumor core had only CA19-9 expression, only sTRA expression, both, or neither (Fig. 2.2E). (If the marker was expressed in  $>1\%$  of the tissue pixels, we counted the core as expressing the marker.) We saw the repeatable occurrence of tumors with predominant sTRA or CA19-9, with good agreement across the TMAs and among individual

sections of the same TMA (Fig. 2.2C). An aberration was the first section of TMA2, which was the first section run and had some poor quality images due to lack of optimization. Also, TMA6 had no cores with only sTRA expression, which we attributed to natural variation between tumors because the images showed no obvious defects in data quality. The two TMAs containing xenograft tumors were similar to the other TMAs, indicating that the expression of each marker persists in culture and in animals. The consistent occurrence of tumors that predominantly express sTRA (Fig. 2.2E), as well as the elevation of sTRA-only regions in the tumors relative to adjacent tissue (Fig. 2.2D), affirm that sTRA is a marker of pancreatic cancer independent of CA19-9.

### **2.3.2 The sTRA and CA19-9 glycans identify spatially and morphologically distinct subsets of cancer cells**

We next explored whether the cells expressing one or the other marker have divergent locations or histomorphologies. The quantification of TMA5 showed that cores were present with various levels of each marker (Fig. 2.3A). Some consistent patterns emerged upon examination of the tissue. In areas of well-differentiated PDAC, CA19-9 staining generally was more prevalent than sTRA (Fig. 2.3B). High CA19-9 in the absence of sTRA also occurred in moderately-to-poorly differentiated PDAC (Figs. 2.3C and 2A.1). Cells that primarily or exclusively expressed sTRA showed other morphologies. A common feature was vacuolated cells<sup>130</sup> (Figs. 2.3B and 2.3D); and less common was sparse, moderately-differentiated glands amidst heavy desmoplasia (Fig. 2.3E). In some cases, the non-invasive ductal epithelium stained mostly with CA19-9 while the invasive cells were strongly positive for sTRA (Fig. 2.3B). Certain tumors showed a subpopulation of sTRA-expressing cells with large cytoplasm (Fig. 2.3F) adjacent to moderately-differentiated glands expressing CA19-9 (Fig. 2.3F). We found that the well-differentiated epithelium with foamy cytoplasm<sup>131</sup> always expressed both sTRA and CA19-9 (Fig. 2.3G and 2A.1).

The two markers, therefore, are present in non-identical subsets of cancer cells. The cancer cells variously express either one, both, or neither of the markers, and the morphologies of the cells group into a few categories in association with the exclusive or dual expression of CA19-9 and sTRA.

We asked if the above observations hold true in model systems. The results from a TMA containing cell-line xenografts and a TMA containing patient-derived xenografts were similar to the results from the primary tumors. Among the 10 cell lines on TMA68, some expressed both markers, others only one marker, and others neither (Fig. 2.4A). The 14 PDX models on TMA69 showed divergent expression of the two markers; eight primarily expressed CA19-9, and six primarily sTRA (Fig. 2.4B). The tumors from the cell-line xenografts generally showed less stroma and ductal epithelium than primary tumors, whereas the PDX models had better recapitulation of the primary tumors. The tumors expressing primarily CA19-9 or sTRA were largely similar to each other in histomorphology (Figs. 2.4C, 2.4D, 2A.2 and 2A.3), although one of the cell lines with exclusive sTRA expression, Panc05.04, showed a lipid-rich phenotype (Fig. 2.4C). We observed this phenotype also in primary tissue with heavy sTRA staining (Fig. 2A.1). These analyses show that the CA19-9 and sTRA expression phenotypes persist in cultured cancer cells and are not just unique to the primary tumors.

### **2.3.3 Protein expression differs based on glycan type and differentiation**

We next asked if the expression of key markers of phenotype are different between the groups identified above. We stained for various protein markers in a separate color from the glycans (Fig. 2.1D), acquiring measurements of three proteins in each of two sections from TMAs 2 and 5. In one section we stained for E-cadherin, vimentin, and cytokeratin 19 (CK19). We found that nearly all cells expressing either CA19-9 or sTRA also expressed CK19 and E-cadherin, regardless of morphology (Fig. 2A.4), and that none expressed vimentin (not shown), indicating a consistent epithelial phenotype.

We saw more diversity between cells in MUC5AC, a marker of neoplastic and mucin-producing glands, and  $\beta$ -catenin, a marker of regeneration or adhesion (Fig. 2.5). We probed for these proteins along with PDX-1 in another section. The well-differentiated PDACs expressing both sTRA and CA19-9 showed membranous expression of  $\beta$ -catenin, but the poorly-differentiated cells generally showed weaker membranous staining, consistent with a loss of epithelial adhesion. The dual-labeled epithelium with clear cytoplasm always expressed MUC5AC, as did well-differentiated neoplastic glands with luminal secretions, but the poorly-differentiated neoplastic cells never expressed MUC5AC. A tumor with exclusive expression of CA19-9 in moderately-differentiated PDAC with no surrounding stroma expressed neither  $\beta$ -catenin nor MUC5AC, but a moderately-differentiated PDAC secreting primarily sTRA expressed both. These analyses provide evidence that the dual-labeled cells often represent cohesive and secretory epithelia, and that the single-labeled cells are more frequently dyshesive and non-secreting, although additional phenotypes occur.

#### **2.3.4 The expression patterns of sTRA and CA19-9 predict time-to-progression**

We next explored whether certain patterns of glycan expression are associated with the rate of disease progression. For one of the TMAs, Pan CA4, we had information for 45 of the 61 patients about the time from surgery to progression of the disease. We constructed Kaplan-Meier curves of time-to-progression (TTP) grouped by low (< median) and high ( $\geq$  median) expression of each of the five markers (Fig. 2.6A). The amount of staining of neither the individual CA19-9 or sTRA markers nor their exclusive expression markers were related to TTP, but the amount of dual expression was related to TTP ( $p = 0.008$  based on log-rank test). The expression of this glycan type was significantly different ( $p = 0.01$  based on Wilcoxon rank-sum test) between patients with long TTP ( $>2$  years) and those with short TTP ( $<2$  years) (Fig. 2.6B). We used the Marker State Space (MSS) method<sup>132</sup> to test if any patterns of glycan expression distinguish tumors with long TTP from those with short TTP. The method revealed that below a

certain threshold in dual-marker expression, most patients (20/23) had short TTP (Fig. 2.6B). Among the patients above the threshold in dual-marker expression, high levels of both CA19-9-only and sTRA-only were present only in patients with short TTP (Fig. 2.6B).

A three-marker panel consisting of CA19-9-only, sTRA-only, and dual-labeled formed a candidate biomarker for predicting TTP. By classifying patients with high expression of all three markers or with low expression of the dual marker as short TTP, we observed 90% sensitivity (27 correct out of 30) and 80% specificity (12 correct out of 15) for predicting short TTP (Fig. 2.6C). When we ran 10-fold cross validation five separate times, the average accuracy of panels from the training sets applied to set-aside samples was 78% (Table 2.1), which is less than the 87% accuracy (39/45 correct) found for the true grouping, as expected, but robust. Random groupings of the 45 samples as cases and controls resulted in no marker panels (except for one occurrence out of 50) that met the minimum performance of 80% sensitivity and 80% specificity in the training sets (Table 2.1). These results—the good performance in the true grouping and the poor performance in the random groupings—support the interpretation that the marker panel is detecting a true difference between short and long TTP.

An examination of the staining patterns and histomorphology also suggested differences between the groups (Figs. 2.6D and 2A.5). The short TTP tumors were of two types: high in all three markers, or low in the dual marker (Fig. 2.6C). The short-TTP tumors that are high in all three markers frequently showed clusters of poorly-differentiated cells expressing either sTRA or CA19-9 and a lack of differentiated glands (Fig. 2.6D, top row), and in other cases, abundant vacuolated cells or densely-populated, well-differentiated glands (Fig. 2A.5). The short-TTP tumors that are low in the dual marker did not have differentiated glands but rather scattered cells expressing one or the other glycan (Fig. 2.6D, middle row). The long-TTP tumors with expression of the dual marker generally showed well-differentiated PDACs with polarized nuclei and sometimes with vacuole-type cytoplasm (Fig. 2.6D, bottom row). The three short-TTP

tumors that were misclassified did not have well-differentiated glands, nor did the misclassified long-TTP tumors (Fig. 2A.6); additional modifications to the biomarker panel may be required for such tumors.

We asked whether the glycan levels associated with location and type of recurrence (distant or local) or presence of SMAD4 staining (the absence of which is a surrogate for SMAD4 genetic deletion). We did not see significant associations among the 12 patients for which we had the information, but we saw suggestive trends, such as higher dual-marker expression in patients with local recurrence, and higher CA19-9-only staining in lesions that are negative for SMAD4 (Table 2A.2).

The protein expression in the cancer cells was similar to the protein expression presented above, but with higher  $\beta$ -catenin (not shown). Differences in protein levels and morphologies may have been induced by the prior treatment with chemotherapy of the tumors in TMA Pan CA4. In aggregate, these analyses provide preliminary indications that the quantification of the exclusive and dual expression of sTRA and CA19-9 could be useful for classifying tumors and for predicting the risk of disease progression.

## **2.4 Discussion**

Here we show that subpopulations of cancer cells in pancreatic adenocarcinomas are distinguishable by whether they express sTRA, CA19-9, or both. Tumors variously displayed one or more of the subpopulations, sometimes more than one in the same tumor. Each subpopulation had its own characteristics of morphology. Cells expressing both markers typically were part of well-differentiated and mucin-secreting PDACs, whereas those expressing just one were often poorly differentiated and vacuolated and never mucin secreting. Evidence that the glycan-defined subpopulations have predictive value comes from the associations with TTP. We found indications that tumors with short TTP come in at least two varieties: one with an absence of dual-labeled cells, and another with high levels of both the CA19-9-only cells and

the sTRA-only cells. Nearly all patients with long TTP had high levels of the dual-labeled cells without high levels of the single-labeled cells. The findings suggest that the dual-labeled glands indicate a functioning and recovering pancreas, while the dispersed, single-labeled cells mark uncontrolled growth and dispersion. The predictive value of the biomarkers must be validated in follow-up research; at this point, the current research confirms that the sTRA and CA19-9 glycans in combination identify separate subpopulations of cancer cells. In the second place, the research provides an initial look at differences between the subpopulations.

An enabling component of the present work was the automated and quantitative analysis of multimarker fluorescence data. The huge number of images generated in this study would have been impossible to analyze manually, and the analysis would have been only semi-quantitative. The method used here allowed quantification of image data from multiple markers and multiple TMAs and ultimately identification of staining patterns that showed associations with outcome. Another important component was the quantification of exclusive and the dual expression of the markers, which provided better classifications than individual measurements. We foresee such a system having usefulness for research and eventually for clinical applications. In clinical applications, automated image analysis could help to remove inter-operator variability or to pick out rare or subtle features. For example, in OGCs, where the malignant type can be a histiocyte-like sarcomatoid carcinoma cell<sup>116</sup>, automated image analysis could find signals from a stain for such cells amidst an overwhelming background of non-malignant cells.

Most of the previous studies aimed at categorizing pancreatic tumors concentrated on protein or genomics markers. An integrated analysis of mutational status, expression profiles, and histopathology found four subtypes of tumors, defined respectively as low exocrine and high squamous differentiation; increased pancreas-specific progenitor programs; high exocrine and/or endocrine features; or increased expression of immune-specific genes<sup>63</sup>. Other research emphasized the epithelial-mesenchymal transition (EMT) as a means of identifying invasive

cells (see reviews<sup>133,134</sup>). The biomarkers of EMT were not highly specific to cancer cells, but immunostaining for SMAD4—a key node in signal transduction relating to EMT and other functions—may be useful for predicting distant metastasis<sup>61</sup>. Further research focused on developmental pathways, stem-cell antigens, markers of acinar-ductal metaplasia, among others<sup>135-137</sup>.

Linking the glycan types found here to the biological programs and genomics classes mentioned above could be important for further defining and understanding the subpopulations. We could not directly link our data to the genomics classes, but some inferences are possible. For example, the squamous subtype found earlier had low pancreatic differentiation and worse prognosis, suggesting a connection to the single-labeled sTRA and CA19-9 cells in the present study, which were often in nests of undifferentiated cells with no glandular differentiation and were associated with short TTP. Secondly, the previously-defined progenitor subtype had high pancreatic differentiation and mucin expression, suggesting a link with the dual-labeled, mucin-secreting cells found here. Such comparisons could facilitate studies of the biology of tumors, but for practical application, distilling the information down to a small number of immunostains may have significant value. The use of a small number of markers limits the possible number of subgroups; cellular stains provide a direct look at minority cancer cells within complex backgrounds; and cell-surface markers open the opportunity for immunological targeting.

Studies of glycans in subpopulations of cancer cells are less common than studies of proteins and nucleic acids, but they have strong foundations. Several glycans are widely used as markers of cell type, including the CD15 antigen for neutrophils<sup>138</sup>, the TRA-1-60 and TRA-1-81 antigens for induced pluripotent stem cells<sup>139</sup>, the ABO antigens for red blood cells, and the target of the *Lycopersicon esculentum* lectin for endothelial cells<sup>140</sup>. Considering that specific glycans frequently have roles in regulating cellular interactions, it follows that the glycans would be remodeled when cells change states. Pancreatic cancer cells exhibit such behavior. Glycans

altered in pancreatic cancer include CA19-9<sup>141</sup>, members of the Lewis blood group family<sup>142</sup>, and ABO blood group antigens<sup>142</sup>. Except for CA19-9, such glycans were not highly specific to cancer cells. In contrast, sTRA appears to be highly specific to cancer cells, at least in the context of pancreatic tumors. A similar glycan that is present on a glycolipid called sialosyllactotetraosylceramide, or LSTa, has been observed in small-cell lung carcinoma<sup>143</sup> and glioma<sup>144</sup>, and its non-sialylated version is a marker of embryonic stem cells and induced pluripotent stem cells<sup>139</sup>. Consistent with expression on newly-differentiating cells, the sTRA antigen is potentially the precursor of the CA19-9 antigen<sup>145</sup>. These factors suggest that a stem-like population expresses the non-sialylated TRA antigen, and that as cells transform into PDACs, they modify the TRA antigen with sialylation and/or fucosylation. The glycans do not have fully characterized functions, but alterations by fucosylation or sialylation can modulate binding with selectins<sup>146</sup> and galectins<sup>147</sup>, which in turn could alter signal transduction, cell differentiation, and cell migration. The functions may depend on the protein or lipid carriers of the glycans, which have not been characterized for sTRA.

A model arising from this work is that particular glycan and protein combinations define unique differentiation states of cancer cells, making it possible to differentiate tumors based on their content of each type of cancer cell. Such markers could be readily applied to cytologic smears from FNA, which can be difficult to interpret by morphology. An area for potential application would be to help determine which patients should have surgery and which should immediately begin drug treatment. If further research shows that distinct subpopulations have differential responses to available therapies, another application is to select therapeutic regimens. Considering the increasing range of drugs available for PDAC, the opportunities are expanding for matching subtypes to their optimal drugs.

## **2.5 Materials and Methods**

### **2.5.1 Tissue samples and tissue microarrays**

The study was conducted under protocols approved by the Institutional Review Boards at the Van Andel Research Institute, the University of Pittsburgh Medical Center, and the Medical University of South Carolina. All subjects provided written, informed consent, and all methods were performed in accordance with the relevant guidelines and regulations. The tissue samples were collected from extra portions of surgical resections for pancreatic cancer. At each site, tissue microarrays were generated from 1 mm cores of formalin-fixed, paraffin-embedded (FFPE) tissue.

### **2.5.2 Multimarker immunofluorescence and chemical staining**

We performed immunofluorescence and chemical stains on 5  $\mu\text{m}$  thick sections cut from formalin-fixed, paraffin-embedded blocks. We removed paraffin, performed antigen retrieval by incubating the slides in citrate buffer at 100°C for 20 minutes, and blocked the slides in 1X phosphate-buffered saline containing 0.05% Tween-20 (PBST0.05) and 3% bovine serum albumin (BSA) for 1 hour at RT. We labeled two primary antibodies respectively with Sulfo-Cyanine5 NHS ester (13320, Lumiprobe) and Sulfo-Cyanine3 NHS ester (11320, Lumiprobe) according to the supplier protocol. Each round of immunofluorescence used two different antibodies, one against a glycan and one against a protein (see Table S3 for details about the antibodies). After dialysis to remove unreacted dye, we prepared a solution containing both antibodies at 10  $\mu\text{g}/\text{mL}$  in PBST0.05 with 3% BSA. We incubated the antibody solution on a tissue section overnight at 4 °C in a humidified chamber.

The next day, we decanted the antibody solution and washed the slide twice for 3 minutes each in PBST0.05% and once for 3 minutes in 1X PBS. We dried the slide by blotting and incubated Hoechst 33258 (1:1000 dilution in 1X PBS) for 10 minutes at RT to stain nuclei. Following two five-minute washes in 1X PBS, we added a coverslip and scanned the slide using a scanning-

fluorescence microscope (Vectra, PerkinElmer). The microscope collected 35 images at each field-of-view, each image at a different emission wavelength.

We stored the slides in a humidified chamber between rounds of immunofluorescence. Prior to the next round, we removed the coverslip by immersing the slide in deionized water at 37 °C for 30-60 minutes, or until the coverslip came off, and quenched the fluorescence using 6% H<sub>2</sub>O<sub>2</sub> in 250 mM sodium bicarbonate (pH 9.5-10) twice for 20 min. each at RT. The subsequent incubations and scanning steps were as described above.

To treat the slide with sialidase, we incubated a 1:200 dilution (from a 50,000 U/mL stock) of the enzyme ( $\alpha$ 2-3,6,8 Neuraminidase, P0720L, New England Biolabs) in 1X enzyme buffer (5 mM CaCl<sub>2</sub>, 50 mM pH 5.5 sodium acetate) overnight at 37 °C. We washed the slides as above prior to the following antibody incubations. The hematoxylin and eosin (H&E) staining followed a standard protocol.

### **2.5.3 Image and data processing**

We used in-house software called SignalFinder to locate pixels containing signal in each image. The program uses our recently-published SFT algorithm<sup>129</sup> without user intervention or adjustment of settings. From the 35 images captured for each region, we selected the three that corresponded to the emission maxima of Hoechst 33258, Cy3, and Cy5. For each image, SignalFinder creates a map of the locations of pixels containing signal and computes the percentage of tissue-containing pixels that have signal. To arrive at a final number for each core, we averaged over all images for a core. To quantify exclusive or colocalized signals between markers, we used in-house software called ColocFinder. The program allows the user to build up expressions of AND, OR, and NOT between scans, and then quantifies the percentage of pixels that fulfill the expression. The AND operator requires signal pixels to be present in both scans, the OR operator requires pixels to be present in either scan, and the NOT operator requires pixels to be present in the first but not the second scan.

We further analyzed and prepared the data using Microsoft Office Excel and GraphPad Pro, and we prepared the figures using Canvas 14 and Canvas Draw (ACD Systems). The SignalFinder and ColocFinder programs are available upon request.

#### **2.5.4 Statistical analysis**

We used Wilcoxon signed rank test to compare the distribution of a biomarker score between paired samples (e.g. tumor tissue versus adjacent tissue). In the presence of multiple biomarkers (CA19-9, CA19-9-only, sTRA, sTRA-only, Dual), false discovery rate was computed using the Benjamini & Hochberg method. We used Wilcoxon rank sum test to compare the distribution of a biomarker score between two independent groups (i.e. patients with short TTP and patients with long TTP). Kaplan-Meier curves were plotted to characterize the distribution of time-to-progression. Log-rank test was conducted to compare TTP distributions between two groups (i.e. patients with low and high marker values).

#### **2.5.5 Biomarker panel selection using MSS**

We selected marker panels using the Marker State Space (MSS) method<sup>132</sup>. The program searches for marker “states,” or patterns of high and low marker values, that are predominant either in cases or controls and that form accurate classification rules. MSS limits the initial size of panels to 3 markers, with the option of adding markers iteratively. The MSS software is available upon request.

The MSS software has the option of 10-fold cross validation. The program randomly divides all samples into 10 groups, sets aside one group as a test set, and runs the marker search process on the remaining samples. It scans through all threshold combinations to find any panels that meet a minimum accuracy (set by the user, here 80% sensitivity and 80% specificity) in the training samples. The program then applies each panel to the set-aside samples to classify each sample as a case or a control. It compares the true states of the set-aside samples to the

classifications made by the panel to determine the accuracy of the classifications. The accuracy is given as the percentage of classifications that were correct. If no panels give the minimum performance in the training set of a particular split, no panels are applied to the set-aside samples, and no accuracies are given. The program repeats the entire process for all 10 splits.

For further testing, we randomly assigned each of the 45 samples as a case or control. That is, instead of using the actual case or control status of each sample as the input to the program, we used a random assignment of case or control status for each sample. If the marker selection method were simply overfitting classifiers to the data, it would find good classifiers regardless of how the samples are grouped. But if real differences exist between the actual groups, the performance of the panels should greatly decrease when the grouping is randomized.

#### **2.5.6 Patient-derived xenograft (PDX) and cell-line xenograft models**

All animal studies were approved by the VARI Institutional Animal Care and Use Committee (IACUC), and all experiments were performed in accordance with relevant guidelines and regulations. The xenograft studies used 6–8 week old mice from the VARI breeding colony.

The tissue for the PDX models was obtained from surgical resections for pancreatic cancer performed at regional hospitals in Grand Rapids, Michigan, under protocols approved by institutional review boards at the respective institutions. Unused portions of the resections selected by the attending pathologist were placed in a sterile receptacle and transported immediately on ice to the VARI. Upon receipt, the tumor tissue for implantation was placed into a sterile dish containing sterile phosphate buffered saline and carefully teased into  $\leq 3$  millimeters (longest axis) tumor fragments.

The original PDX models were developed in athymic nu/nu mice as reported earlier<sup>148</sup>.

Dependent on tumor tissue availability, tumor fragments were implanted in a maximum of five mice. Mice for each PDX model were gender matched to the donor patient. Following

administration of general anaesthesia (isoflurane), the right flank was cleaned with 70% ethyl alcohol, a small incision made, and a subcutaneous pocket created by blunt dissection. The tumor fragment was inserted into the pocket and the incision closed using a surgical staple. Immediately following surgery, the mouse received a single dose of the analgesic Ketoprofen (5 mg/kg body weight). Mice were monitored for health and tumor growth for the duration of the study, and body weights were recorded weekly. Tumorgraft volumes ( $\frac{1}{2}$  x length x depth x height) were measured 1x/week when volumes  $\leq 50$  mm<sup>3</sup> and 3x/week at tumor volume  $> 50$  mm<sup>3</sup>. A tumorgraft model that failed to develop within 6 months in the 1st generation mice was discontinued and the mice euthanized. When a tumorgraft reached a volume of  $\geq 1500$  mm<sup>3</sup> the mouse was euthanized, and the tumorgraft was aseptically harvested.

For the PDX tumors used for this study, cryopreserved PDX fragments were thawed rapidly, rinsed in sterile phosphate buffered saline containing 1% penicillin/streptomycin (Invitrogen), and implanted into NSG mice. Following administration of general anesthesia (isoflurane), the right flank was cleaned with 70% ethyl alcohol, a small incision made, and a subcutaneous pocket created by blunt dissection. The tumor fragment was inserted into the pocket and the incision closed using a surgical staple. Immediately following surgery, the mouse received a single dose of the analgesic Ketoprofen (5 mg/kg body weight). The monitoring and harvesting were as described above. A portion of the tumor was fixed and processed in a standard manner for histological analysis and TMA construction.

To develop cell-line xenograft models, athymic nu/nu mice were injected with  $1 \times 10^6$  cells/100  $\mu$ L phosphate buffered saline in their right flank using a 1 cc syringe with a 27 g needle. The cell lines were obtained from the American Type Culture Collection (Manassas, VA) and grown in recommended conditions prior to subcutaneous injection. The rest of the methods were identical to those described for the patient-derived xenograft models.

## **2.6 Acknowledgements**

This work was supported by the National Cancer Institute (Alliance of Glycobiologists for Cancer Detection, 1U01CA168896; Early Detection Research Network, U01CA152653; R21CA207779 and R21CA186799); the National Institute of General Medical Sciences (1R41GM112750); the National Institute for Allergy and Infectious Disease (R21AI129872); and the Biorepository and Tissue Analysis Shared Resource, Hollings Cancer Center, Medical University of South Carolina (P30CA138313).

We thank the VARI Pathology and Biorepository Core for assistance with preparing sections and TMAs from the xenograft samples; Dr. David Monsma and the VARI Animal Services Core for preparation of the mouse xenografts; the VARI Confocal Microscopy & Quantitative Imaging Core for assistance with high-throughput fluorescence scanning; and Dr. Edward Zhou for assistance with glycan structural rendering.

# Chapter 3: The sTRA Plasma Biomarker: Blinded Validation of Improved Accuracy over CA19-9 in Pancreatic Cancer Diagnosis

Ben Staal,<sup>1\*</sup> Ying Liu,<sup>1\*</sup> Daniel Barnett,<sup>1,2\*</sup> Peter Hsueh,<sup>1</sup> Zonglin He,<sup>3</sup> ChongFeng Gao,<sup>1</sup> Katie Partyka,<sup>1</sup> Mark W. Hurd,<sup>6</sup> Aatur D. Singhi,<sup>4</sup> Richard R. Drake,<sup>5</sup> Ying Huang,<sup>3</sup> Anirban Maitra<sup>6</sup>, Randall E. Brand,<sup>4</sup> Brian B. Haab<sup>1</sup>

<sup>1</sup>The Van Andel Research Institute, Grand Rapids, MI

<sup>2</sup>Michigan State University, East Lansing, MI

<sup>3</sup>Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>4</sup>University of Pittsburgh Medical Center, Pittsburgh, PA

<sup>5</sup>Medical University of South Carolina, Charleston, SC

<sup>6</sup>MD Anderson Cancer Center, Houston, TX

\*Equal contributions

Running title: The sTRA Plasma Biomarker

Keywords: Pancreatic cancer, plasma biomarker, glycans, surveillance, diagnosis

Financial support

NCI: U01 CA152653 (BH, RB, YH); U01 CA200466 (RB); U01 CA200468 (AM); U01 CA168896 (BH, RB, YH); U01 CA196403 (AM); P30 CA138313 (RD).

Conflict of interest: The authors declare no potential conflicts of interest.

Correspondence:

Brian B. Haab, PhD

Van Andel Research Institute

333 Bostwick NE, Grand Rapids, MI 49503

[brian.haab@vai.org](mailto:brian.haab@vai.org)

Tel. (616) 234-5268

Word count: (3324 max); Figure count: 5; Table count: 1.

Significance: No serological biomarkers have been validated to improve upon CA19-9 for pancreatic cancer diagnostics. We show that a glycan called sTRA forms that basis for a biomarker that could meet clinical requirements for surveillance or differential diagnosis of pancreatic cancer.

Accepted for Publication in Clinical Cancer Research 21 December 2018

### **3.1 Translational Relevance**

Here we report a new biomarker for pancreatic cancer, called sTRA, that yields better performance than CA19-9, the current best biomarker for pancreatic cancer. sTRA is produced by pancreatic cancers that do not produce CA19-9. As a result, biomarker panels including sTRA gave improved specificity or sensitivity. In a rigorous, double-blinded study, the panels performed well enough to potentially warrant clinical use. One panel could be valuable for surveillance for incipient pancreatic cancer among people with elevated risk, and another panel could be valuable for differential diagnosis relative to benign pancreatic disease. Such biomarkers could lead to improved outcomes for many patients afflicted with pancreatic cancer.

### 3.2 Abstract

*Purpose.* The CA19-9 biomarker is elevated in a substantial group of patients with pancreatic ductal adenocarcinoma (PDAC), but not enough to be reliable for the detection or diagnosis of the disease. We hypothesized that a glycan called sTRA is a biomarker for PDAC that improves upon CA19-9. *Experimental Design.* We examined sTRA and CA19-9 expression and secretion in panels of cell lines, patient-derived xenografts, and primary tumors. We developed candidate biomarkers from sTRA and CA19-9 in a training set of 147 plasma samples and used the panels to make case/control calls, based on predetermined thresholds, in a 50-sample validation set and a blinded, 147-sample test set. *Results.* The sTRA glycan was produced and secreted by pancreatic tumors and models that did not produce and secrete CA19-9. Two biomarker panels improved upon CA19-9 in the training set, one optimized for specificity, which included CA19-9 and two versions of the sTRA assay, and another optimized for sensitivity, which included two sTRA assays. Both panels achieved statistical improvement ( $p < 0.001$ ) over CA19-9 in the validation set, and the specificity-optimized panel achieved statistical improvement ( $p < 0.001$ ) in the blinded set: 95% specificity and 54% sensitivity (75% accuracy), compared to 97%/30% (65% accuracy). Unblinding produced further improvements and revealed independent, complementary contributions from each marker. *Conclusions.* sTRA is a validated serological biomarker of PDAC that yields improved performance over CA19-9. The new panels may enable surveillance for PDAC among people with elevated risk, or improved differential diagnosis among patients with suspected pancreatic cancer.

### 3.3 Introduction

The proper management and treatment of cancer begins with reliable detection and diagnosis of the disease. Reliable detection and diagnosis can be particularly challenging for pancreatic ductal adenocarcinoma (PDAC), owing to the internal location of the tumors, similarities to benign conditions, and heterogeneity between patients in the makeup of the tumors. A molecular feature shared by most PDACs is increased levels of a glycan called the CA19-9 antigen. CA19-9 is used for specific purposes, such as to confirm the diagnosis of PDAC, assess responses to treatment, or screen for recurrence, but it has limitations<sup>149-151</sup>. It is not useful for the substantial group of patients without elevations in the marker, and it shows a ~25% false-positive rate among patients with benign conditions of the pancreas using a threshold that gives a ~75% true-positive rate<sup>152</sup>. Elevated cutoffs provide <5% false-positive rates, but with detection of just 25-50% of patients<sup>149</sup>. CA19-9 by itself, therefore, is not sufficient for rendering a diagnosis or for unequivocally assessing responses to treatment. On the other hand, it detects a major subset of patients and is still one of the most-used biomarkers in oncology. In fact, over the several decades since the discovery of CA19-9, no biomarker has been established to surpass its performance.

We previously investigated the concept that the tumors that do not overproduce CA19-9 are different from those that do, and that they produce alternate glycans that are structurally similar to the CA19-9 antigen. One class of glycans we found is based on a structural isomer of the CA19-9 antigen called sialyl-Lewis X<sup>153,154</sup>. The sialyl-Lewis X glycan showed elevations in 30-50% of the patients with low CA19-9 but also showed elevations in about 10% of patients with benign pancreatic diseases. Another glycan, referred to as sTRA, was elevated in up to half of the patients with low CA19-9, with very low false-positive rates<sup>127</sup>. In subsequent research, we found that the cells producing sTRA are different in location, morphologies, and molecular characteristics than the cells producing CA19-9<sup>155</sup>. The above findings suggested that the sTRA glycan would be a serological biomarker for pancreatic cancer that could improve upon CA19-9.

Many previous studies have examined candidate biomarkers for PDAC (see reviews <sup>156-158</sup> and discussion). Based on information from the previous work, we incorporated several considerations into this study. The most rigorous test of a biomarker is to apply it to independent, blinded samples, make case/control calls on each sample, and assess performance by comparing the calls to a “true” case/control status based on a gold standard. Most reports of candidate biomarkers do not include such a test. In this study, the gold standard was the diagnosis arrived at through the full information available for each patient, and a benchmark was the performance of CA19-9. We further ensured a rigorous test of performance by emphasizing the detection of resectable cancer (stage I/II cancers), and by testing specificity for cancer relative to benign conditions of the pancreas.

Another unique aspect of this study is an examination of the biomarker production and secretion in tumor models and primary tumors. The most effective cancer markers are the ones produced and secreted by the cancer cells, rather than as secondary effects from the liver or inflammatory processes. An analysis of biomarker production across tumor models and primary tumors, together with an assessment of the secreted levels in each, could help to confirm that the biomarker is directly produced by the cancer cells and that elevations in the blood plasma result from secretion by the cancer cells. Such a study also could confirm the complementary relationship between CA19-9 and sTRA, that many cancers that do not produce CA19-9 produce sTRA.

In this study, we demonstrate that sTRA provided significantly improved performance over CA19-9 in a double-blinded test using preset thresholds and classification rules. The improved performance was the result of complementary elevations among CA19-9 and two versions of the sTRA assay, comprising a three-marker panel. Studies of cell-culture and patient-derived xenograft models of pancreatic cancer and primary tumors confirmed these relationships.

## **3.4 Methods**

### **3.4.1 Human specimens**

The study was conducted under protocols approved by the Institutional Review Boards at the Van Andel Research Institute, the University of Pittsburgh Medical Center, MD Anderson Cancer Center, the Mayo Clinic, and the Medical University of South Carolina. All subjects provided written, informed consent, and all methods were performed in accordance with an assurance filed with and approved by the U.S. Department of Health and Human Services. All collections took place prior to any surgical, diagnostic, or medical procedures. The donors consisted of patients with pancreatic cancer or a benign condition involving the pancreas, and from healthy subjects (Table 1). The healthy subjects had no evidence of pancreatic, biliary or liver disease. All blood samples (EDTA plasma) were collected according to the standard operating procedure from the Early Detection Research Network and were frozen at -70 °C or colder within 4 hours of time of collection. Aliquots were shipped on dry ice and thawed no more than three times prior to analysis.

### **3.4.2 Sandwich immunoassays**

The antibody array methods followed those presented earlier<sup>159-161</sup> with slight modifications. The capture antibodies were CA19-9 (1116-NS-19-9, MyBioSource), anti-MUC5AC (45M1, Thermo Scientific), and anti-MUC16 (X325, Abcam). The biotinylated primary antibodies were CA19-9 (clone 1116-NS-19-9, MyBioSource) or TRA-1-60 (TRA-160, Novus Biologicals). The secondary detection agent was Cy5-conjugated streptavidin (Roche Applied Science). The Supplementary Methods contain details of the assays, the calibrators and controls, and the processing of biomarker data and the acquisition of immunofluorescence data.

### **3.4.3 Statistical methods**

The case/control comparisons of individual biomarker values measured on a continuous scale were performed using the two-sided Student's t-test. The case/control comparisons of gender used the Fisher's Exact test, and the comparisons of age used the Wilcoxon rank-sum test. To

assess relationship between biomarkers and covariates, we presented Spearman correlation between biomarker and continuous covariates and tested for equivalence in biomarker distribution across covariate categories using Wilcoxon Rank Sum test (when there are two categories) or the Kruskal-Wallis rank sum test (when there are more than two categories). To test for difference in the average of sensitivity and specificity between a panel and CA19-9, we computed bootstrap standard error of the summary measure using nonparametric bootstrap<sup>162</sup> with 1000 resamples stratified on case/control status, and computed two-sided p-value with Wald test. Statistical analyses were performed using R statistical software (version 3.5.1).

### **3.5 Results**

#### **3.5.1 Detecting the sTRA and CA19-9 glycans**

The CA19-9 antigen (Fig. 3.1A) is a tetrasaccharide detected by the CA19-9 monoclonal antibody<sup>163</sup>. A monoclonal antibody called TRA-1-60<sup>128</sup> detects the presumed precursor of the CA19-9 antigen, a non-fucosylated and non-sialylated tetrasaccharide<sup>108</sup> (Fig. 3.1A). In order to indirectly detect the sialylated version of the TRA-1-60 antigen, which is referred to as sTRA (sialylated TRA), we treat the antigen with sialidase prior to detection (Fig. 3.1A). Both CA19-9 and sTRA appear on multiple glycoproteins and glycolipids<sup>164,165</sup>. In the blood of pancreatic cancer patients, we previously detected the glycans primarily on the mucins MUC1, MUC5AC, and MUC16, and more rarely on MUC5B and MUC3A<sup>127,160,166</sup>. We further showed that the cancer cells producing CA19-9 are separate from those producing sTRA<sup>155</sup>. If the cancer cells secrete the antigens accordingly (Fig. 3.1B), we would expect plasma samples to show elevations of one, both, or neither of the markers with frequencies similar to observed in tissue. The standard CA19-9 assay uses a CA19-9 antibody for both capture and detection (Fig. 3.1C). For sTRA, we detected the antigen on three different capture antibodies: CA19-9, anti-MUC5AC, and anti-MUC16 (Fig. 3.1C). The combinations of capture and detection antibodies are referred to as CA19-9:sTRA, MUC5AC:sTRA, and MUC16:sTRA, respectively.

### **3.5.2 The sTRA antigen in CA19-9-negative cancer models and primary tumors**

To determine whether various models of pancreatic cancer make and secrete sTRA, and whether it is produced by some that do not produce CA19-9, we examined a panel of 10 cell lines derived from pancreatic cancers. Some of the cell lines produced only CA19-9, others only sTRA, and others both or neither (Fig. 3.2A & 3.2B). The amount secreted into the media roughly corresponded to the amount on the cell surfaces (Fig. 3.2B and Fig. 3A.1), and certain cell lines secreted almost exclusively only one of the glycans (Fig. 3.2B). Patient-derived xenograft (PDX) models potentially provide a more faithful representation of primary tumors. Across a panel of 13 PDX models, sTRA was produced and secreted by several tumors showing low levels of CA19-9 (Figs. 3.2C & 3.2D), and the levels of sTRA and CA19-9 in the sera correlated with tumor expression (Fig. 3A.1). The prevalence of each type could be different from those observed in clinical plasma samples, since differences could exist between the types in the take rates in culture or in PDX mice, but the models confirm that some PDACs make only one of the glycans, and others make both.

Next, we used a tissue microarray to determine glycan expression in the primary tumors of 52 patients, and we used the CA19-9 and sTRA sandwich assays (Fig. 3.1D) to determine the levels in matched blood plasma. The staining in the tumors was diverse (Fig. 3.3A & 3.3B), as observed in the cell lines and PDX models, and the levels in blood plasma showed that certain patients had elevations in only CA19-9 or sTRA (Fig. 3.3B). The blood levels of each marker correlated with the tissue levels (Fig. 3A.1). Overall, the models and primary tumors showed that sTRA is produced by a substantial subset of PDACs, that the secreted levels reflect the tumor levels, and that it occurs in many cases not showing production or secretion of CA19-9.

### **3.5.3 Improved classification performance using the combined markers**

To explore the performance of sTRA as a plasma biomarker, we measured CA19-9 and the three sTRA assays (complete data in Table 3A.1) in an initial set of blood plasma from 147 subjects (Table 3.1). As an individual marker, the CA19-9:sTRA assay performed similarly to

CA19-9 (Fig. 3.4A). The CA19-9 performance was in agreement with previous reports<sup>149</sup> and our previous studies<sup>152</sup> on similar cohorts, yielding 70-75% sensitivity at 70-75% specificity (Fig. 3.4A). All markers except MUC5AC:sTRA had significantly-higher ( $p < 0.05$ ) means in stage III-IV than in stage I-II cancer (Table 3A.2), but the overall biomarker performance, as assessed by receiver-operator characteristic analysis, was only slightly higher in stage III-IV cancer (Fig. 3A.2). None of the markers showed a significant difference between control types (Table 3A.2). The relationships between the assays were the same as in the model systems—complementary, non-correlated elevations in the sTRA and CA19-9 assays (Fig. 3.4B). We therefore sought to develop a biomarker panel that included any combination of CA19-9 and the sTRA assays. Using the MSS method<sup>132</sup>, we identified two lead panels, one that provided high specificity for the detection of cancer (low false positive rate), and another with high sensitivity (low false negative rate). A threshold is applied to each of 2 or 3 markers, and each pattern of is assigned as a “case state” or a “control state” (Fig. 3.4C, and Supplementary Methods for details on the thresholds used for each marker). By classifying the subjects with an elevation in any member of the panel as a case, overall performance was improved relative to CA19-9, both for the specificity-optimized panel and for the sensitivity-optimized panel that did not include CA19-9 (Fig. 3.4D).

We then applied the biomarker panels to independent samples, comprising 25 cases and 25 controls with similar makeup as the training set (Table 3.1). We used the predetermined thresholds and classification rules from the 147-sample training set to make a case/control call on each sample (complete data in Table 3A.1). For CA19-9, the thresholds also were based on the training set—one to give high specificity, and another to give high sensitivity—and subjects with levels above the threshold were called as cases. The increases in average sensitivity and specificity over CA19-9 was statistically significant for both panels ( $p < 0.001$ , 1000-fold bootstrapping), and improvements in either sensitivity or specificity relative to CA19-9 were consistent with the training set (Fig. 3.4D).

In both panels, we saw that a substantial percentage of patients were in the complementary subsets of patients that were classified as cases (Fig. 3.4E), indicating that each member of the panels contributed independent information. In addition, the complementary contributions of the individual panel members were consistent between the training and validation sets.

#### **3.5.4 Blinded validation of improved sensitivity and specificity**

We then applied each panel to a new set of 147 samples that was run blinded. We applied the predetermined thresholds, made a case/control call on each sample, and sent the calls to a separate site for determination of performance. The predetermined thresholds for both the panel biomarkers and CA19-9 were based on the combined 147-sample training and the 50-sample validation sets. The data and thresholded results are in Tables 3A.3 and 3A.4.

The panel optimized for specificity gave high specificity and improved sensitivity over CA19-9 from 30% to 54%. The panel optimized for sensitivity gave moderate gains over CA19-9 in both sensitivity and specificity (Fig. 3.5A). The difference in the average of specificity and sensitivity was statistically significant ( $p < 0.001$ ) for the specificity panel, and the difference was positive but not statistically significant ( $p = 0.18$ ) for the sensitivity panel (Fig 3.5A).

The performance of the individual panel members and their relationship to each other was consistent with the training and validation sets. The individual CA19-9:sTRA assay performed similarly to CA19-9 and better than the other sTRA assays (Fig. 3.5B), and complementary elevations were observed between CA19-9 and the sTRA assay (Fig. 3.5C). The CA19-9 and CA19-9:sTRA assays were correlated, due to two samples with high levels in both, but several samples were elevated in only one or the other of the assays. The marker levels were higher ( $p < 0.05$ ) in stage III-IV cancers (Table 3A.2), but the AUCs in ROC analysis were similar between stage I-II and stage III-IV cancers (Fig. 3A.2). Among the controls, benign biliary stricture and chronic pancreatitis showed higher levels than the other control groups in CA19-9 and CA19-9:sTRA (Table 3A.2). Such elevations are commonly observed, and the difference from the training set is likely due to natural variation.

Because the training set may not fully represent the whole population of cases and controls, we investigated whether a simple adjustment of the individual marker thresholds would improve the performance of the panels or CA19-9. The adjusted specificity-optimized panel gave 96% specificity and 65% sensitivity, better than the optimized CA19-9 performance of 96% specificity and 46% sensitivity (Fig. 3.5D). The adjusted sensitivity-optimized panel gave 96% sensitivity and 37% specificity, but CA19-9 gave just 9% specificity at 96% sensitivity (Fig. 3.5D). The improvements of the panels relative to CA19-9 were very similar between the test set and the full 197-sample training set.

In both the test set and the full, 197-sample training set, each member of the panels provided independent, complementary value (Fig. 3.5E). The percentages in patient subsets were remarkably similar between the sets. These results indicate that the relationships between the individual markers were consistent over all sets, and that the marker panels gave consistently improved performance over CA19-9.

### **3.6 Discussion**

A biomarker that improves upon CA19-9 would be a significant advance in diagnostics for pancreatic cancer, given the fact that no biomarker has achieved that feat in the several decades since the development of CA19-9. The uses for such a biomarker could include screening or surveillance for pancreatic cancer, and differential diagnosis of pancreatic cancer relative to benign conditions. Whether a new biomarker will find value in clinical application depends on the performance requirements of the application.

For the early detection of pancreatic cancer, screening among the general population is not viable because the prevalence of the disease is too low to justify the cost. An alternative strategy is surveillance for incipient pancreatic cancer among a population with elevated risk. An elevated-risk condition that has gained attention in recent years is sudden-onset type-2 diabetes<sup>167</sup>. In that group, the prevalence of pancreatic cancer may be as high as 0.8%<sup>168</sup>. At such a prevalence, a biomarker with 96% specificity and 65% sensitivity would have a positive

predictive value (PPV) of 11.6% and negative predictive value (NPV) of 99.7%, which could be acceptable in a cost-benefit analysis. Thus, the biomarker panel presented here is achieving the performance required for use in surveillance among elevated risk groups.

For differential diagnosis, the goal is to differentiate cancer from non-cancer among people with a suspected abnormality of the pancreas, for example individuals with abnormal imaging of the pancreas in an initial evaluation. In the application of a blood test to such patients, those positive for the test could receive further workup or treatment, and those negative for the test could be spared unnecessary procedures, thus reducing cost, risk, and emotional burden to the patient. In this use of a blood test, high sensitivity is critical. The prevalence of pancreatic cancer among referral patients with abnormal imaging would vary greatly between centers, but it could be as high as 15% (the experience of the collaborators in the present study). As such prevalence, a biomarker with 96% sensitivity and 50% specificity would have  $PPV = 25.3\%$  and  $NPV = 98.6\%$ , potentially high enough to find adoption.

Other serological biomarkers have shown promise for the diagnosis of pancreatic cancer and will be important for comparative studies. Many have been investigated <sup>156-158</sup>, more than can be listed individually, but the following are some important examples. Plasma thrombospondin-2 was combined with CA19-9 to yield high specificity and sensitivity in multiple sample sets <sup>169</sup>, and a drop in specific isoforms of apolipoprotein AII strongly discriminated pancreatic cancer from healthy controls, although not from benign diseases, in a blinded study <sup>170</sup>. Panels of biomarkers including metabolic markers <sup>171</sup> and protein indicators of a migratory signature <sup>172</sup> showed particularly encouraging results in recent studies. One of the most promising developments has been the detection of mutated, cell-free DNA in the circulation of cancer patients. The great majority of patients with pancreatic cancer harbor oncogenic mutations in the KRAS genes in their tumors. A PCR-based assay to detect such mutated DNA in the circulation identified about 30% of pancreatic cancer patients with near-perfect specificity relative to healthy controls, and the combination with CA19-9 and other markers could increase

sensitivity to 64% at 99.5% specificity<sup>173</sup>. The generalization of this strategy to include additional mutations showed promise for screening for eight common cancer types, including pancreatic cancer<sup>174</sup>. Further research will address specificity among benign conditions and performance in blinded studies. The performance of the panels in this study compares favorably with those cited above, and the precise, relative merits could be determined in comparison studies using common samples. Given that combining CA19-9 with the PCR-based assay improved sensitivity<sup>173</sup>, it is reasonable that the addition of sTRA would further improve sensitivity.

The present study has certain limitations. The samples were collected prior to knowledge of diagnosis, which is one of the PROBE design requirements<sup>175</sup>, but they were not collected in prospective manner that mimicked clinical application. The training and validation sets included cases and controls all collected from the same location and same setting, but in the test set, some controls were collected at a separate site to include subjects with diabetes. For further validation, the sample size should be expanded; prospective sample collection at multiple sites should be used; and the measurements should be acquired using the clinical assay that would be used in practice<sup>176</sup>.

The overall performance of the panels potentially could be improved through additional glycans in the Lewis blood group, of which CA19-9 is a member called sialyl-Lewis A (sLeA). Some pancreatic cancers have upregulated tumor expression of an isomer of sLeA called sialyl Lewis X (sLeX)<sup>142</sup>, which we<sup>153,154</sup> and others<sup>177</sup> found elevated in the circulation of many pancreatic cancer patients. Other patients elevate a glycan detected by the DUPAN-2 monoclonal antibody<sup>178,179</sup>, identified primarily as type 1 sialyl-LacNAc<sup>180,181</sup>. The elevation of CA19-9 in the blood potentially results from accumulations in the stroma followed by leakage into the capillaries or lymph<sup>182,183</sup>. Therefore, new leads potentially could be found by analyzing tumors with a non-glandular histopathology using glycan-discovery methods such as whole-tissue MALDI imaging<sup>184</sup>.

This research establishes the sTRA glycan as a new biomarker for PDAC that improves diagnostic accuracy over CA19-9. This is the first biomarker, to our knowledge, to statistically-significantly improve upon CA19-9 in a double-blinded test with preset thresholds and classification rules. The applicability of the findings to future PDAC samples is supported by the similar breakdowns of distinct, complementary groups in each set and the similar improvements in performance between sets. Furthermore, the importance of sTRA was supported by its expression and secretion in pancreatic cancer models and primary tumors that do not produce CA19-9. The true value will become clearer over time, but at this point it appears the new biomarker identifies a distinct subset of PDACs. Based on the performance observed here, the biomarker panels could be valuable for surveillance among elevated-risk people or for the differential diagnosis of pancreatic cancer.

### **3.7 Acknowledgements**

We thank the VARI Confocal Microscopy and Quantitative Imaging core for assistance with fluorescence image acquisition on the tissue samples; and Luke Wisniewski at VARI for assistance preparing the cell cultures.

## Chapter 4: sTRA and CA19-9 expression distinguish independent pancreatic tumor cell subpopulations and prognosis

## 4.1 Abstract

Subpopulations of cancer cells that express specific and necessary characteristics are likely to be predisposed to facilitate or initiate invasion and metastasis. Identification of these cells could provide a critical tool for studying and targeting metastasis in pancreatic cancer. We have previously characterized the sTRA glycan and CA19-9 antigen (sialyl-Lewis A glycan) as independent subpopulations of pancreatic ductal adenocarcinoma (PDAC) cells. We hypothesized that the expression of these glycans identifies subpopulations of PDAC cells with predisposition to invasion and metastasis in pancreatic tissues. The glycans sTRA and CA19-9 as well as their co-expression and independent exclusive expression are validated in 5 tissue microarrays (TMAs) with 64 matched tumor and adjacent tissues. Tumors showed significant ( $p < 0.001$ ) elevation over adjacent uninvolved tissues in all glycan expression groups. The three glyco-type states, sTRA-only, CA19-9-only and their co-expression (“dual”) are correlated between tumor and lymph node and tumor and metastases. There were strong Pearson correlations for CA19-9 (0.6685) and dual (0.7134) in lymph nodes, and moderate correlation (0.4724 and 0.4650) in lymph nodes and metastases. Patients with long-term overall survival had high dual expression. Validation of survival thresholds applied to overall survival had a positive predictive value of 86.2% and negative predictive value of 50% for short-term survival. Together, glycan-expressing subpopulations of pancreatic cancer may show prognostic differences for metastasis and survival that could be exploited for future study and intervention.

## 4.2 Introduction

In part due to such high resistance to therapy as well as poor early diagnosis and high recurrence rates post-resection, pancreatic cancer is now the third leading cause of cancer mortality in the United States with 8% survival at 5 years.<sup>185</sup> Multiple mechanisms of chemotherapy and other treatment resistance have been proposed.<sup>186-188</sup> Strong evidence has suggested stromal density is part of the resistance to chemotherapies and indeed it has been demonstrated that chemotherapy penetration of tumors can be poor,<sup>186</sup> but removal of stroma actually resulted in worse clinical outcomes.<sup>189</sup> Other studies have suggested a role for PD-1, CTLA4 and other immunotherapy targets,<sup>190</sup> but clinical trials also showed equivocal response.<sup>95</sup> Thus far, many potential therapies have shown promise in animal studies, but none have shown high success in humans and some have even resulted in poorer outcomes from treatment. The standard of care therapies are FOLFIRINOX and Gemcitabine/nab-paclitaxel with best survival of 11.4-13.8 months and 9.8-12.1 months with average duration to treatment failure of 4.3 and 3.7 months.<sup>3,4</sup>

In pancreatic cancers and particularly in pancreatic ductal adenocarcinomas (PDAC), tumors tend to have low neoplastic cellularity with significant implication for biological activity of pancreatic tumors.<sup>76</sup> Within those limited neoplastic cells there is high clonal heterogeneity that also show significant biological and clinical implications for treatment and survival.<sup>8,191</sup> Further, these subpopulations tend to have varied abilities to become quiescent or activate as environmental conditions dictate to evade harsh conditions present in PDAC tumors and death by treatments.<sup>7,192</sup>

Even without consideration for the also heterogeneous stroma of PDAC tumors, the characteristics of neoplastic subpopulations vary widely and discovery of biomarkers to identify the most aggressive neoplastic cell subpopulations could provide stratification for treatment and outcomes-based decisions. Further, discovery of biomarkers to identify high risk vs low risk

subpopulations could also provide insight into the biology of cells and allow isolation or categorization for further study and treatment development.

Most neoplastic subpopulation studies to date have examined DNA, RNA, and epigenetic modifications for tumor population partitioning and/or subtype stratification.<sup>8,63,76,89,94,193</sup> Other studies have identified populations of cells by gross immunohistochemical staining of individual protein markers in tissue.<sup>36,61</sup> We have previously shown that subpopulations can also be separated by glycan expression state.<sup>155,194</sup> Specifically, we demonstrated differences in tissue and serum expression of CA19-9 (sLeA glycan) and its near relative glycan, sTRA in pancreatic cancer.

When dually-expressed, these glycans together potentially suggest a baseline cancerous phenotype with longer survival of patients, while expression of CA19-9 alone or sTRA alone potentially show shorter survival.<sup>155</sup> Further, we previously demonstrated that some tumors expressing only CA19-9 have glycan expression in patient tumors without detectable secretion to blood plasma.<sup>194</sup> Meanwhile sTRA was detected in blood plasma of all patients expressing sTRA in tissue<sup>194</sup>. Although only some carriers of these glycans have been identified (e.g. mucins, like MUC1, MUC3, MUC5AC, MUC16),<sup>195</sup> the variable expression of these glycans suggests differences in release and trafficking of sTRA and CA19-9 from and on cells in pancreatic tumors.

We also preliminarily demonstrated the use of the three biomarkers (dual, CA19-9 only, and sTRA only) as a biomarker panel for tissue to assess risk for short- or long-survival in neoadjuvant-treated patients. This suggests a potential application to develop subtypes to more accurately determine clinical patient risk of progression based on the presence of neoplastic cell subpopulations.

In this work, we hypothesize that neoplastic pancreatic cells co-expressing sTRA and CA19-9 form a subpopulation of pancreatic cancer cells that represent a baseline less aggressive glycotype and cell subpopulations that lose expression of one glycan represent a more

aggressive pancreatic cancer glyco-type presenting higher risk for metastasis and changes in resistance to treatment. Therefore, each of the three subpopulations of neoplastic cells (Dual, CA19-9 only, and sTRA only) represents an independent cell subpopulation, each with its own implications for clinical progression and response.

We previously presented a method of multimarker immunofluorescence with quantification by the SignalFinderIF tissue immunofluorescence analysis software.<sup>129,155,196</sup> Here, we utilized this method to validate the separation of glyco-type subpopulations on additional tissue microarrays. We then evaluated the expression of the glycotypes in tumors against their matched lymph nodes and metastasis samples to evaluate metastatic risk of tumor glycotypes. We then compared these results to survival of patient groups to validate our previous findings.

## **4.3 Materials and Methods**

### **4.3.1 Multimarker Immunofluorescence (MMIF)**

The multimarker immunofluorescence method has been previously described.<sup>155</sup> In brief, antibodies are labeled with Cy5 or Cy3 for each round of immunofluorescence staining. All antibodies to be applied in any round are tested for cross reactivity prior to use. Formalin fixed and paraffin embedded (FFPE) tissue is deparaffinized and rehydrated and antigen retrieval is performed by 30 minute incubation at 100°C in citrate buffer. Slides are washed with 1X phosphate buffered saline (PBS) then incubated with Cy3- and Cy5-labeled antibodies for 2 hours at room temperature or overnight at 4°C. Slides are incubated with Hoechst dye for 10 mins and a coverslip applied with aqueous mounting medium. Slides are imaged on a Vectra2 (Perkin Elmer, Hopkinton, MA) fluorescence microscope at 20X (0.5um pixel resolution). The slide is incubated overnight in a humidified chamber then incubated in milli-Q or equivalent water at 37°C for 10-20 minutes to remove the coverslip. The slide is quenched with 6% hydrogen peroxide in 250mM sodium bicarbonate (pH 9.5) for 20 minutes, twice. After quench successive rounds of antibody application, imaging and quenching are repeated until all antibodies have been run. For sTRA detection, pan-neuraminidase is applied to slides at 37°C

overnight and then detected with the TRA-1-60 antibody (Novus Biologicals, Littleton, CO) labeled as above.

#### **4.3.2 SignalFinderIF Software Analysis**

SignalFinder-IF and its algorithm have been previously described.<sup>129,196</sup> Quantification is performed with the Segment Fit Thresholding (SFT) algorithm and provides robust signal determination by analyzing local background to threshold signal. Quantified images are then analyzed by ColocFinder<sup>196</sup> with 0.5 inclusion and 5 pixel box size for colocalization (e.g. glycan1 AND glycan2) or exclusion (e.g. glycan1 NOT glycan2) parameters.

#### **4.3.3 Tissue Microarrays**

Tissue microarrays were assembled by the University of Pittsburgh Medical Center (UPMC), the Medical University of South Carolina (MUSC), and the University of Nebraska Medical Center (UNMC). Survival data was collected by UPMC. All human samples are used in compliance with Institutional Review Board protocols approved by the University of Pittsburgh Medical Center, Medical University of South Carolina, University of Nebraska Medical Center and the Van Andel Institute. Written and informed consent was provided by all subjects. The UPMC samples were collected from surgical resection tissue and assembled in 1mm punch cores (1-2 punches per patient per tissue). The MUSC samples were collected from surgical resections with matching adjacent uninvolved (“Normal”) tissue and lymph node biopsies or as autopsy samples with matching tumor, adjacent uninvolved tissues, and metastases. The UNMC sample collection was through the Rapid Autopsy Program (RAP). They were assembled as 5mm punch cores and include matching tumor and metastatic tissues collected within 2-6 hours of death.

#### **4.3.4 Statistical Analysis**

Tumors, tumor lymph nodes and metastases were tested for normality and log-normalized, then normalized to a 0-1 scale by marker by tissue prior to correlation. Correlations were tested by Pearson’s correlation. Linear regressions were performed by ordinary least squares and the

95% confidence interval displayed using R software (R-project, Vienna, Austria). Significance testing was performed by Wilcoxon Rank Sum for all non-normalized data. Kaplan-Meier curves were graphed for Time-To-Progression (TTP) or Overall Survival (OS) analysis and log-rank test was used to test for significance between marker groups.

## **4.4 Results**

### **4.4.1 Validation of sTRA and CA19-9 elevations as independent indicators of PDAC**

CA19-9 antibodies have been well-characterized as antibodies that detect the sLeA glycan (i.e. type-1 LacNAc with terminal sialic acid and fucose modifications, Figure 4.1A).<sup>197</sup> sTRA is an antigen detected indirectly by applying pan-neuraminidase ( $\alpha$ -2-3,6,8) to tissues then detecting with the TRA-1-60 antibody. TRA-1-60 is an antibody that has been characterized as detecting a unique moiety consisting of a terminal type-1 LacNAc  $\beta$ 1-3 linked to a type-2 LacNAc with the antibody having a very high selectivity against the sialylated variant (sTRA) and high specificity for the N-acetyl on the glucose on the glycan root<sup>108</sup>. The sTRA glycan has near relative glycans, such as LSTa ending in a glucose at the base of the glycan, often found as a glycolipid,<sup>108,179</sup> and the likely glycan detected by DUPAN2 (sialylated type-1 LacNAc),<sup>179</sup> but none have shown the specificity and sensitivity we have demonstrated for discrimination of pancreatic cancer with sTRA and CA19-9.<sup>194</sup>

Our previous work showed significant elevation of sTRA, CA19-9 and a combination of the two glycans in tumors on three clinical tissue microarrays (TMAs) including matched adjacent uninvolved tissue.<sup>155</sup> Prior to pursuing further analysis, we sought to validate the significance of the biomarker findings in additional clinical TMAs of matched PDAC tumor and adjacent uninvolved tissues. In Figure 4.2B, there was significant variation between TMAs, which may indicate natural clinical variation or variable staining and imaging conditions. Despite this, there were significant positive outliers or distributions trending toward significant differences from adjacent tissues on every TMA, even with the relatively small n on each individual TMA (Figure 4.2B). In aggregate (Figure 4.2C), the glycan expression levels were much more significant. All

five groups showed significant elevation ( $P < 0.001$  or lower) in tumor against matched adjacent uninvolved tissue, demonstrating utility in sTRA as an adjunct marker to CA19-9, as well as the exclusive subpopulations (Dual, CA19-9-only, and sTRA-only). It also serves as a validation of the previous comparisons performed on the smaller TMA set.

#### **4.4.2 Glycan expression in tumors are correlated with distant recurrence**

Glycan expression between tumors, lymph nodes, and metastases may indicate whether glycan expression is an inherent condition of a clonal population of cells or induced by environment. Either outcome may be exploited for diagnostic state or treatment effect, but the answer has significant implications for future studies. Prior to this study, sTRA expression had not been characterized in lymph nodes, although the presence of CA19-9 is typically negative in lymphatic tissues. We show sTRA is negative in normal lymph nodes on both TMAs with significant differences in tumor lymph nodes and normal lymph nodes for all markers (Figure 4.3A). The expression of all glycans in all tissues with neoplastic cells over their normal matched tissue further validates sTRA as a biomarker of neoplastic pancreatic cancer cells. We then examined the relationship between primary tumor expression and tumor lymph node glycan expression and found strong Pearson correlation values within the CA19-9-only (0.6685) and Dual (0.7134) expression subpopulations. Between primary tumors and metastases, primary tumor expression of sTRA-only showed moderate correlations with sTRA-only (0.4650) in distant metastases as well as dual (0.5297) expression in distant metastases (Figure 4.3B). With higher correlation values between tumors and lymph node metastases for CA19-9-only and dual expression, this suggests that CA19-9 may play a role in entry of tumor cells into lymphatics for malignant pancreatic cancer. The role of CA19-9 for lymphatic metastases is bolstered by moderate correlation between sTRA-only in tumor with dual expression in lymph nodes as these populations rise in CA19-9 expression. It is also notable that neither CA19-9-only nor sTRA-only expression was correlated with a direct switch to the other solo-expressing glycotype in lymph node metastases. CA19-9-only was weakly correlated with dual expression

in tumor lymph nodes suggesting that sTRA is not a critical factor in cancer progression to lymph nodes.

Meanwhile, sTRA-only was the highest correlation among the glycotypes for more distant metastases in Set 1 (MUSC), though the correlation was moderate (0.4724) (Figure 4.3C). The lack of correlation between solo glycan subpopulations was observed again with weak or no correlation in tumor sTRA-only against metastatic CA19-9-only as well as in the converse correlation. Only weak to moderate correlations were observed from tumor dual expression to metastasis, indicating that dual expressing tumors may be similarly likely to switch from dual to a solo glycan expressing glycoctype in metastasis.

To test these findings, we obtained an independent sample set (Set 2) from the University of Nebraska Medical Center Rapid Autopsy Program (RAP). In this small sample set, the metastases showed correlations similar to tumor lymph nodes with moderate correlation within CA19-9-only and a strong correlation for dual staining in metastases from both dual and sTRA-only expressers (Figure 4.3D). Further validation will be needed to confirm these relationships, but in this first assessment in matched tumors with regional and distant metastases, subpopulation glycotypes in tumors appear to be associated with glycotypes expressed in metastases. This indicates glycan expression characteristics are retained from primary tumor to metastasis or restored after migration.

#### **4.4.3 High dual expression and high solo expression confer differences in survival in clinical tumors**

We previously showed high dual CA19-9/sTRA co-expression (upper 50%) associates with longer progression-free survival. In the present study, we reanalyzed 42 of those subjects (UPMC1) in addition to 43 new subjects (UPMC2) for overall survival. None of the markers tested to significance in a binary analysis, but dual expression again showed trended differences in longer survival with high expression (Figure 4.4B). Upon further review, the upper 15% of dual marker expression show a significant difference (longer survival) from the

remaining 85% (data not shown), though the significance of this finding will have to be validated in future sample sets.

The samples were thresholded by analyzing quantified signal for combinations of CA19-9, sTRA, CA19-9-only, sTRA-only, and dual expression by three signal thresholding cutoffs ( $3\sigma$  and  $6\sigma$ ). Thresholds for percent signal were determined by the Marker State Space (MSS) software, which optimizes thresholds and biomarkers for biomarker panels<sup>132</sup>. The thresholds were applied to the TMAs and combined data set and class states determined for long- and short- survivors. The same markers were determined to be optimum (sTRA- $6\sigma$ , CA19-9-only -  $3\sigma$ , and Dual- $6\sigma$ ) in two independent optimization runs on the two UPMC TMAs, though with varied thresholds. The thresholds were averaged and applied to both runs. The analysis showed the panel had a specificity of 60% and a sensitivity of 80.6% with a negative predictive value of 50% and positive predictive value of 86.2%. Although the performance did not replicate the previous high performance (90% specificity and 80% sensitivity) on progression-free survival, there may be room to for further improvements to obtain higher positive predictive value for better clinical guidance. Kaplan Meir plots did not show significant differences in log-rank testing (Figure 4.4C). However, when the 8 potential combinations of the three markers were analyzed for contribution to survival prognosis, they showed two significantly different groups (Figure 4.4D). The sTRA+/Dual+ group showed significantly longer survival and the CA19-9 only+/sTRA+ group showed significantly higher risk for short survival. This result will need to be further validated in future data sets.

## **4.5 Discussion**

We showed here that subpopulations of pancreatic ductal adenocarcinomas identified by the glycotypes CA19-9-only, sTRA-only, and dual expression of both CA19-9 and sTRA differentiate subpopulations of cells in terms of aggressiveness (i.e. propensity to metastasize or propagate). Both regional and distant metastases frequently expressed glycotypes similar to primary tumors.

In cases where metastases expressed different glycotypes from their associated primary tumors, the glyctype most frequently changed from dual expression in tumors to solo marker expression in metastases, indicating a loss of a single glycan expression. The more frequent single-glycan expression in metastasis from dual-glycan expression in primary tumors suggests potential benefit to PDAC cell subpopulations with solo expression in metastatic formation and trafficking. The role of dual expression as a less aggressive state for PDAC is also supported by the moderate association of dual expression in metastases with all three glycotypes. Further, the short survival time and time to progression for low dual-expressing tumors also supports dual expression as a glyctype of lower aggression for tumors and a possible marker for better prognosis.

In this study we attempted to reach a prognostic biomarker based on overall survival, which is ultimately a more difficult, although more meaningful endpoint for all cancers. The lack of separation by most of the glycotypes may demonstrate that due to the short survival of the vast majority of patients, these glycotypes may be better suited for prognosis of progression rather than overall survival. As biomarkers of progression, they could provide utility, though more limited than a successful biomarker with long-term stratification, but it could be an indicator of biology predisposition for metastasis. Nonetheless, the significant differences in the sTRA+/CA19-9-only+ and sTRA+/Dual+ group further supports the hypothesis that dual expression represents a less aggressive tumor phenotype than solo expression of CA19-9. In our previous studies, we showed that CA19-9-only and sTRA-only were associated with poorer differentiation states in primary tumors<sup>155</sup>. This may be an indication of loss of epithelial cell type and potential transition to mesenchymal cell features of epithelial to mesenchymal transition (EMT). The cells that make this transition and seed metastases are thought to constitute a neoplastic cell subpopulation representing 1-5% of cancer cells and are known as metastasis-initiating cells (MICs).<sup>198</sup> Once these cells re-establish in a pre-metastatic niche in a distant metastatic site, they recruit myofibroblasts to establish a new stromal compartment to

support the growth of the cancer cells and transition to a more epithelial phenotype.<sup>199</sup>

Metastases may re-establish dual expression as a return to epithelial state.

If the differences in glycan expression between lymph node metastasis and distant metastasis suggested here are validated in future datasets, this could define a differential role for CA19-9 over sTRA in trafficking to lymph nodes. The sialyl Lewis A (sLeA) antigen recognized by CA19-9 is known to bind E- and L-selectins,<sup>200</sup> of which L-selectins are abundantly present in lymph nodes.<sup>201</sup> Selectin binding by CA19-9 in hematogenous spread of pancreatic cancer has been demonstrated previously.<sup>202</sup> Selectins have been shown to bind sialylated and fucosylated lactosamines,<sup>203</sup> of which sLeA is one though sTRA is not, and may represent a differential route to metastasis for CA19-9. At this time, there are no known binding targets to sTRA. Metastasis to lymph nodes is very frequent even in small and resectable PDAC tumors and have been shown occur around the same rate as CA19-9 in patients (65-90%).<sup>204-206</sup> Standard histopathology may also miss tumor lymph nodes for the small micrometastases and tumor cell subpopulations that establish in lymph nodes.<sup>204,206</sup> A CA19-9-only trafficking route may be further leveraged for diagnosis. We have previously demonstrated that some tumors expressed CA19-9-only without secretion to blood plasma with worse patient outcomes.<sup>194</sup> With lymphatic trafficking following a different mechanism, it is possible that CA19-9 could be detectable in lymph fluid despite the lack of blood plasma secretion.

Risk stratification by subpopulation glycoform could allow therapies to be targeted to higher risk cell subpopulations, whether metastasis has already occurred or not. Preliminary data for drug sensitivity to first-line treatments showed greater sensitivity for CA19-9-expressing cells (data not shown), suggesting that while they are more likely to follow some routes of metastasis, their populations may be more treatable with current therapies, if therapies can reach their targets. Future studies should consider testing for glycoform and determine therapies that can target these more resistant cells.

More, the analyses here add to the validity of the three glycotypes as subpopulations of neoplastic cells in pancreatic tumors. We demonstrate a significant difference between tumor and adjacent in two additional tissue microarrays as well as between tumor and normal lymph nodes. We further establish that sTRA is not natively expressed in normal lymph nodes. sTRA expression has not previously been described in other tissues, although LSTa with a similar terminal glycan structure has been shown in glioma<sup>144</sup> and small cell lung carcinoma,<sup>143</sup> which may denote a common regulatory pathway for glycosylation or functional benefit to the terminal glycan structure for cancers.

Pancreatic cancer heterogeneity has been well characterized with significant implications for its ability to adapt and evolve to survive harsh environmental conditions. It is likely that not all pancreatic cells will be predisposed to survive these conditions or to progress the cancer. However, it is likely that some cell subpopulations have a differential benefit in escaping the primary tumor and traffic to new metastatic sites. A model proposed here is that those cell subpopulations can be discriminated by glycotype. Further, most of these glycotypes can be detected by blood test, though cytologic smears from fine needle aspiration biopsies could also allow more accurate diagnosis of glycotype for the highest risk groups. Once the glycotype is established, it is retained in metastasis and may be used to determine susceptibility to treatment. Alternatively, future efforts could potentially develop treatments to target these cell subpopulations.

## Chapter 5: Conclusions and Future directions

## 5.1 Summary

Defining subpopulations of pancreatic cancer cells allows for the segmentation of cells to characterize pancreatic cancer biology, to better identify biomarkers for diagnosis and prognosis, and to stratify tumors by risk of clinical progression. Pancreatic cancer and particularly pancreatic ductal adenocarcinomas (PDAC) are very heterogeneous, both between and within tumors. PDAC tumors usually have very low neoplastic cellularity (1-20%),<sup>76,77</sup> and of the cells present, they are extremely polyclonal.<sup>8</sup> Although tumor cell populations do seem to follow a progressive evolution,<sup>60</sup> clones from multiple phases of progression have potential to become or be deadly by differing mechanisms (e.g. locally invasive, widely metastatic, neuroendocrine signaling, physically obstructive).<sup>8,60,61,207,208</sup>

Sequencing of RNA, DNA, and miRNA can provide valuable information on clonal populations, particularly with laser-capture microdissection and newer single-cell sequencing methods, but complex sequencing with clinical validity has a long turnaround time. Due to the rapid progression of pancreatic cancer and lack of treatment differences between the primary driver mutations, sequencing often requires too much time to wait for initial treatments.[ref Aguirre APA 2018] Typically, sequencing does not provide information fast enough for first-line treatment and often not for second-line treatment in pancreatic cancer, as well. Sequencing also fails to provide spatial information to determine the distribution of cells in tumors, which is biologically significant for understanding mechanisms of dissemination of disease in a tumor. It is possible to perform sequential genome-wide or exome sequencing on several microdissected samples from a tumor, but these methods have not been widely demonstrated in research or clinical settings and would be extremely costly. Thus, there is a need to identify other methods of determining cell states and subpopulations. Ideally, cell states and subpopulations would be identified by low cost, reproducible methods with biomarkers that could examine available samples and determine diagnosis, prognosis, and stratification of treatments to effectively treat PDAC.

Glycans and glycoproteins are potentially well suited to this purpose. Glycans and glycan-carrying glycoproteins are often displayed on the surface of cells and are frequently secreted to blood, making them available as diagnostic markers. More, many glycans serve functional roles in signaling, transport, and inter-cellular communication and identification, so they may serve dual roles as both a cell subtype identification molecule and a functional target for treatment. With retained presence on cells, glycans can also present a display target for antibody targeting by new types of treatments currently in development (e.g. monoclonal antibody treatment, lipid and nanoparticle drug delivery).

In the preceding three chapters, this dissertation explored three aims to investigate how sTRA and CA19-9 may be used as biomarkers of cell subpopulations with relevance to the diagnosis and prognosis of PDAC. In the first aim, sTRA and CA19-9 expression was examined in PDAC tissues to determine whether they represented subpopulations of neoplastic cells. In the second aim, the effect of sTRA and CA19-9 tissue expression on blood plasma presence and diagnostic power were determined. In the third aim, differences in sTRA and CA19-9 expression were tested for their indication of subpopulation aggressiveness by assessing their likelihood to result in metastatic retention and patient survival.

### **5.1.1 Definition of Cell Subpopulations**

In the first aim of this dissertation, sTRA and CA19-9 were hypothesized to be two different subpopulations of pancreatic ductal adenocarcinoma (PDAC) cells. On analysis, there were actually three subpopulations of PDAC cells defined by glycan expression: those only expressing sTRA (sTRA only); those only expressing CA19-9, CA19-9 only); and those expressing both (Dual expression). The pathological phenotypes of the subpopulations showed spatial and morphological separation. sTRA-only populations associated with poorer to moderate differentiation and foamy cytoplasm. Although sTRA was found on the cell surface in dual expression, it was usually cytoplasmic as a solo expression marker. sTRA was also correlated with an increase in  $\beta$ -catenin on cell membranes, indicating a progressing, though

stabilizing and intact cell adhesion between cells of the ductal membrane. CA19-9-only subpopulations were present in poor to well differentiated populations that ranged from isolated cells in heavy stroma to hyperglandular features. CA19-9 was found both in dual expression with sTRA and expressed alone on cell membranes, though was rarely found in cytoplasm. Dual sTRA/CA19-9-expressing subpopulations were moderately- to well-differentiated with high nuclear polarity and pre- to early- neoplastic features. Dual expression was predominantly found on both intact membranes and budding or embolic features in ductal lumens. Further, these features were consistent across an initial set of clinical tumors as well as patient-derived xenograft and cell line xenograft model systems, indicating the stable presence of the glycans as biomarkers of subpopulations. Thus, based on difference in spatial location, histomorphological traits, and expression of certain proteins, we concluded that pancreatic cancers can be divided into subpopulations according to the three groups defined above.

### **5.1.2 Plasma sTRA and CA19-9 in the Diagnosis of PDAC**

We further defined sTRA as a new biomarker for PDAC with sTRA increasing the diagnostic accuracy of CA19-9, when present. The significant value of sTRA for clinical translation is that it is secreted into the blood plasma. Given that pancreatic biopsies are costly, inconvenient to the patient, physically and emotionally burdensome, and risky, serological biomarkers have huge practical advantages. In the second aim, we investigated the relationships between cellular expression and secreted levels. We showed the association of glycan expression in tissue with distinct secretion patterns to blood plasma. sTRA expressing tissues were consistently associated with sTRA in blood plasma, suggesting sTRA expressing tissues always secrete sTRA to plasma. In CA19-9 expressing tissues, a subset of CA19-9 expressing tissues without sTRA expression (“CA19-9-only” tissues) were associated with no CA19-9 present in plasma, suggesting the CA19-9 glycan failed to be secreted to blood plasma.

In non-secreting CA19-9 expressing tissues, they demonstrated either hyperglandular or blind duct features and scattered cells in dense stromal tissue. In the hyperglandular tissues, there

was a complex network of duct structures that form with relatively little stroma and no organization to the duct network that resemble previously described blind duct structures.<sup>209</sup> We further showed that cell cultures recapitulated the finding of CA19-9 secretion failure to culture media in a subset of PDAC cell lines. Thus, the plasma levels of the glycans are generally good indicators of the tissue phenotype, but a subset of CA19-9-positive (and sTRA-negative) tumors do not secrete CA19-9, potentially indicating a distinct subset of PDAC. With that knowledge, we next investigated the value of sTRA as a diagnostic biomarker. We showed sTRA represents not only an independent subpopulation of cells, but an independent biomarker for diagnosis with equivalent performance to CA19-9 and together a stronger biomarker for diagnosis than either sTRA or CA19-9 alone. This aim of the dissertation established the potential for clinical translation of the sTRA biomarker.

### **5.1.3 sTRA and CA19-9 in PDAC Dissemination**

An important remaining question was whether the distinct subpopulations are different from one another in their behaviors. The combined sTRA and CA19-9 biomarkers improve diagnostic accuracy, but can they also identify differences in aggressiveness or outcome? Subsequently in the third aim, the use of these glycan subpopulations as measures of relative aggressiveness was demonstrated by analysis of glycan retention in metastases and glycan expression correlated with survival of pancreatic cancer cell subpopulations. The TMA data from the first aim was validated with additional TMAs. Lymph nodes showed a significant elevation in all of the glycan biomarker subpopulations in tumor over normal lymph nodes. Metastases showed similar expression of glycans to their matched primary tumors. Interestingly, there was no indication of glycan switching between primary tumors and lymph nodes/metastases (e.g. no correlation in sTRA-only in primary tumors to CA19-9-only in lymph nodes), further validating their representation of independent cell subpopulations and indicating a lack of type-switching upon dissemination.

Glycotypes also varied in their indication of long- and short- survival of patients. We found that the composition of the cells in the tumor was an important variable, rather than the presence or absence of any single glycotype. Tumors with single-expressing (sTRA-only or CA19-9-only) in the absence of any other type was associated with poor prognosis, as were tumors with high levels of all three types (Dual, sTRA-only, and CA19-9-only). In contrast, tumors containing Dual-expressing cells either alone or with only one of the single-expressing types represent a milder phenotype. The sTRA/dual tissue glycotype represents a more epithelial phenotype and less likelihood of mobility to distant sites. It also likely represents a stable growth environment with intact early progression features with few mesenchymal characteristics expected for cell migration. Meanwhile, CA19-9 expression in the absence of sTRA or with cells also expressing sTRA without dual expressing cells are likely to result in higher mobility and metastasis. As shown in the first aim, the expression pattern is found in flatter cells with poor nuclear polarity indicating a more mesenchymal phenotype and higher association with expression in metastases. Thus, this aim established that the glycotypes are stable and that they have differences in metastatic characteristics, with notable differences between the Dual-expressing and the single-expressing cells.

Together, the developments of these three aims make a strong case for the use of glycans in the definition of subpopulations of pancreatic cancer cells and particularly the utility of sTRA and CA19-9 as defining characteristics of independent subpopulations of PDAC tumors.

#### **5.1.4 Development and Application of a Multimarker Quantitative Pathology System**

Further, these studies have advanced the methodology for studying marker expression in tissue. The studies described above were enabled by the development of a multimarker immunofluorescence assay, automated microscopy workflow, and automated software analysis that has increased the objectivity of pathological staining quantification and throughput for enough samples to reach statistical and biological significance. In Barnett, Hall, and Haab (Appendix A), we describe a new software platform and enabling tools that provide flexible,

objective, and high-throughput analysis tools to perform signal quantification, colocalization, and overlay tools for the analysis of immunofluorescence staining with equivalent or better quality performance to previously described tools. These methods will provide additional benefit to the field for future quantification of pathological staining.

These findings open several questions for future analysis. Although the validation of these glycotypes as subpopulations of PDAC cells is strong, the mechanisms by which the differences of these cells' actions are carried out are unclear. Two major areas to explore are the conditions that result in CA19-9 or sTRA secretion to blood plasma and the transit of cells expressing CA19-9 or sTRA for metastasis. These are likely related to each other based on trafficking of proteins or cells by endothelial and lymphatic receptor expression.

## **5.2 Future Directions**

### **5.2.1 Glycan secretion and trafficking**

Previous data suggests that CA19-9-expressing cell migration and metastasis is likely supported by selectins.<sup>200</sup> This is both a target for biologic study and treatment development in cells expressing CA19-9.<sup>210</sup> Binding partners of sTRA are completely unclear. However, from the plasma-tissue correlation study in the second aim,<sup>194</sup> sTRA was more consistently detected in blood plasma, suggesting better secretion of sTRA from tissue to blood plasma. This also correlates with the strong relationship between sTRA-only expression in primary tumors and metastasis to liver, which likely occurs by hematogenous spread.

Future studies should be directed at identifying potential receptors for sTRA in distant tissues and on endothelial cells. One potential study could test dye- or radio-labeled sTRA carriers onto endothelial cell layers to determine if sTRA binds to endothelial receptors as has been used for trafficking studies of CA19-9 and fucose.<sup>211</sup> sTRA could also be used as a ligand for immunoprecipitation of endothelial and liver lysates to attempt pulldown of receptors that could be identified by mass spectrometry, as has been previously demonstrated for CA19-9.<sup>212</sup> sTRA

receptor identification would both contribute to the understanding of mechanism, but also potentially be used to reduce migration with inhibition, if sTRA directly contributes to trafficking. Trafficking and identification of tumors expressing sTRA or CA19-9 could also follow previous positron emission tomography studies in people, similar to preliminary studies performed with CA19-9.<sup>213</sup> In addition to using anti-glycan antibodies, similar studies could be performed with labeled CA19-9 and sTRA glycans, which could identify the location of specific glycan binding ligands and suggest further tissues for pathological analysis. It could also suggest potential routes for migration and metastasis for further characterization and inhibition. There is strong evidence that pre-metastatic niche formation primes potential metastatic sites for establishment of metastases when metastatic emboli find appropriate conditions.<sup>82,214</sup> As ligands, glycans are well known to have a role in trafficking and binding. The presence of sLeA and other lewis glycans have been shown in liver and other tissues.<sup>215,216</sup> With glycans present, it would follow that binding sites for these glycans may also be present. Identification of binding sites in potential metastatic sites for sTRA and CA19-9 could suggest a direct role for these glycans and the subpopulations of cells expressing them for establishment of metastasis.

### **5.2.2 Improved diagnostics by antibody development and additional glycan biomarker discovery**

The validation of sTRA as a viable biomarker and significant contributor to diagnostic accuracy suggests that other glycans may be able to help close the gap to accurate detection of all pancreatic cancers. sTRA represents non-fucosylated sLeC from the Lewis synthesis pathway. This is only one half of the Lewis synthesis tree where ST3GAL transfers sialic acid to LeC (LacNAC)-containing terminal structures. On the other half of the pathway, FUT2 adds an  $\alpha$ 2 fucose to the terminal galactose preventing the addition of sialic acid. This produces increased H group and potentially Lewis b(LeB), when FUT3 is present. These represent two additional glycans that could be powerful biomarkers. More, sTRA and CA19-9 represent accurate detection of 85-90% of pancreatic cancer samples,<sup>155,194</sup> and the secretor phenotype has been

estimated to be present in up to 20% of patients.<sup>101,103</sup> This suggests that the remaining portion of these sTRA and CA19-9 low patients may be shunting additional glycan production to H group and LeB glycans, analogous to sTRA and CA19-9 in the patients that are negative for both CA19-9 and sTRA. This suggests that LeB and H group could identify the remaining double negative patients. Further, as has been shown for separate subpopulations being represented by sTRA and CA19-9 detections, these groups with different glycan expression patterns could further represent additional subpopulations of pancreatic cancer cells with yet different biological characteristics.

Although the  $\alpha$ 1-2 fucosylated H group is present in people with O blood type, H group is a general description for the  $\alpha$ 1-2 fucose attached to terminal galactose of a LacNAc group. sLeA and sTRA are on terminal type 1 LacNAc groups (Gal $\beta$ 1-3GlcNAc). If H group predominant patients have H group or the further fucosylated LeB from FUT2 secretion, they would be produced on this same Type I LacNAc backbone, which is not endogenous in blood group expression<sup>217</sup> in blood. Further, H, A, and B blood groups have been shown to be expressed in pancreatic acinar cells natively, while sLeA has been shown to have native expression to ducts and centroacinar cells.<sup>215</sup> Aside from secretor/non-secretor status, which may or may not have a role in pancreatic secretion,<sup>215</sup> the H group, A, B and LeB expression may be due to the origin of the originating cancer cell (ductal vs acinar origin). The expression patterns of both Lewis glycans and blood group antigens in pancreas may also indicate differentiation and developmental state of the cell.<sup>216</sup> These correlations are further indications that the alternate fate for type I LacNAcs may present valuable markers to strengthen the value of CA19-9 and sTRA as diagnostic biomarkers of pancreatic cancer as well as biomarkers of subpopulations with differential biological determination.

There is also further room to improve the diagnostic accuracy of sTRA. Currently, sTRA is being detected by an indirect detection due to lack of a direct antibody. The development of an antibody to sTRA is likely to give increased detection specificity and sensitivity. In the previous

chapters, we have shown how increasing the number of capture antibodies increased the specificity and sensitivity of glycan detection. This suggests that there are still additional sTRA carriers that are not being detected and the maximum sensitivity may be achieved by a direct sTRA capture and detection antibody.

### **5.2.3 Potential Therapeutic Applications**

There are two potential mechanisms for treatment improvements utilizing glycan subpopulation expression: predicting resistance and susceptibility to current treatments and novel targeted therapeutics. Resistance and susceptibility to current treatments may vary by glycan-expressing subpopulations. Preliminary cell culture data from our lab has shown potential susceptibility to current first-line treatments for CA19-9-only expressing cell lines and resistance to treatments for sTRA-expressing lines. If validated in animal models, this may provide more effective selection of treatment for a subset of pancreatic tumors. Alternatively, a clinical trial could be designed for patients with late-stage cancers where treatment could be selected by glyco-type, if tissue biopsies could safely be obtained. This could potentially be used to determine first-line chemotherapy treatments.

sTRA and CA19-9 may also be used for therapeutic targeting by antibody-linked treatments due to their cell surface expression. Treatments have been developed targeting CA19-9 antigen by antibody integration on nanoparticles and attached to liposomes for chemotherapeutic drug delivery.<sup>210,211</sup> The same could be done with an sTRA targeting antibody to better target more of the total patient population and dual targeting could be used in dual-expressing patients with a likely improved efficacy over CA19-9-targeting trials. In addition to nanoparticle-directed treatments, monoclonal antibodies to sTRA could be used alone or with monoclonal antibodies to CA19-9 as potential direct treatments to pancreatic cancers, similar to CA19-9 strategies tested in animal models.<sup>218</sup> With retention of glycan expression on distant metastases, this could present a viable strategy to target extra-pancreatic metastases.

Glycotype-directed or targeted therapy could provide a new method for timely and effective therapy. More, these therapies could potentially be guided by glycan expression detected in blood plasma given the determined associations from tissue-plasma correlation in chapter 3, though the non-secreting CA19-9-only group would need new diagnostics to track efficacy. The model systems characterized in chapter 2 could be used for preliminary testing of antibody treatments and glycotype-determined chemotherapy. Further validation could then be conducted in additional organoid models.<sup>219</sup> As an extension of preliminary treatment models currently in development, this would provide a much more robust application of these treatment modalities and represents significant potential for clinical impact of glycotype determination on clinical practice.

### **5.3 Concluding Remarks**

Defining subpopulations of pancreatic cancer cells could provide a powerful mechanism for increased understanding of pancreatic cancer and allow for development of better, more targeted and effective therapies for a disease with very high need. Glycans represent a rapidly assessable and highly informative class of molecules for the characterization of pancreatic cancer cell subpopulations. The contributions of the work here show that sTRA and CA19-9 expression can identify unique cell subpopulations of pancreatic cancer cells with differential morphology and differentiation states. More, these features are retained in metastatic states. With differential expression of glycans in metastases, these cell subpopulations also suggest differences in prognosis for pancreatic cancer patients and could potentially provide valuable targets for therapy, especially in currently untreatable metastatic disease. The use of sTRA and CA19-9 for improved diagnostic accuracy provides an improvement to current diagnosis and the biology of glycans suggests there may be additional room for improvement with characterization of further glycans in the sTRA/CA19-9 synthesis pathway for Lewis antigens. Further characterization of sTRA may also provide further insights into the action of glycans in trafficking to blood vessels, if a binding target can be identified. In addition to these biological

contributions, the work here was facilitated by the development of efficient, precise and accurate quantification methods for immunofluorescence in tissues. These tools have significant potential to allow new rapid diagnostic and prognostic assessment of tissues for both research and clinical pathologists with objective and sensitive detection of disease by quantitative pathology. With these considerations, this thesis provides both valuable new contributions to the field of pancreatic cancer biomarkers and has opened new questions for future study.

## APPENDIX

Chapter 1 Tables

Table 1.1 Incidence and survival of selected common cancers

Primary Site	<b>Pancreas</b>	Breast	Prostate	Lung	Colon	Ovarian
Incidence (per 100k)	<b>12.79</b>	131.1	105.00	50.99	26.1	11.49
Cases (total)	<b>61,860</b>	343,965	345,915	264,275	144,128	34,279
-Localized	<b>10%</b>	62%	78%	16%	38%	15%
-Regional	<b>29%</b>	31%	12%	22%	36%	20%
-Distant	<b>52%</b>	6%	5%	57%	23%	59%
-Unknown	<b>8%</b>	2%	4%	5%	4%	6%
Survival (5 year)	<b>9.4%</b>	89.7%	97.7%	19.0%	63.0%	48.3%
-Localized	<b>34.3%</b>	98.7%	100%	56.3%	90.4%	92.3%
-Regional	<b>11.5%</b>	85.3%	100%	29.7%	71.4%	74.5%
-Distant	<b>2.7%</b>	27.0%	30%	4.7%	13.5%	29.2%
-Unknown	<b>5.5%</b>	54.5%	80.9%	7.8%	26.2%	24.8%

All statistics collected from SEER Cancer Statistics<sup>9</sup>

Table 1.2 Cancer screening test performance for tests in clinical use

Cancer ( <i>Screening Test</i> )	Specificity	Sensitivity
Prostate Cancer		
<i>Digital Rectal Exam</i> <sup>11,12</sup>	40-90.7%	28.6-81%
<i>Prostate Specific Antigen (PSA)</i> <sup>14</sup>	93.8%	20.5%
Breast Cancer		
<i>Mammogram</i> <sup>20</sup>	90.5-92.5%	83.2-97.9%
Lung Cancer		
<i>Low dose computed tomography (LDCT)</i> <sup>#28</sup>	28-100%	80-100%
Ovarian Cancer		

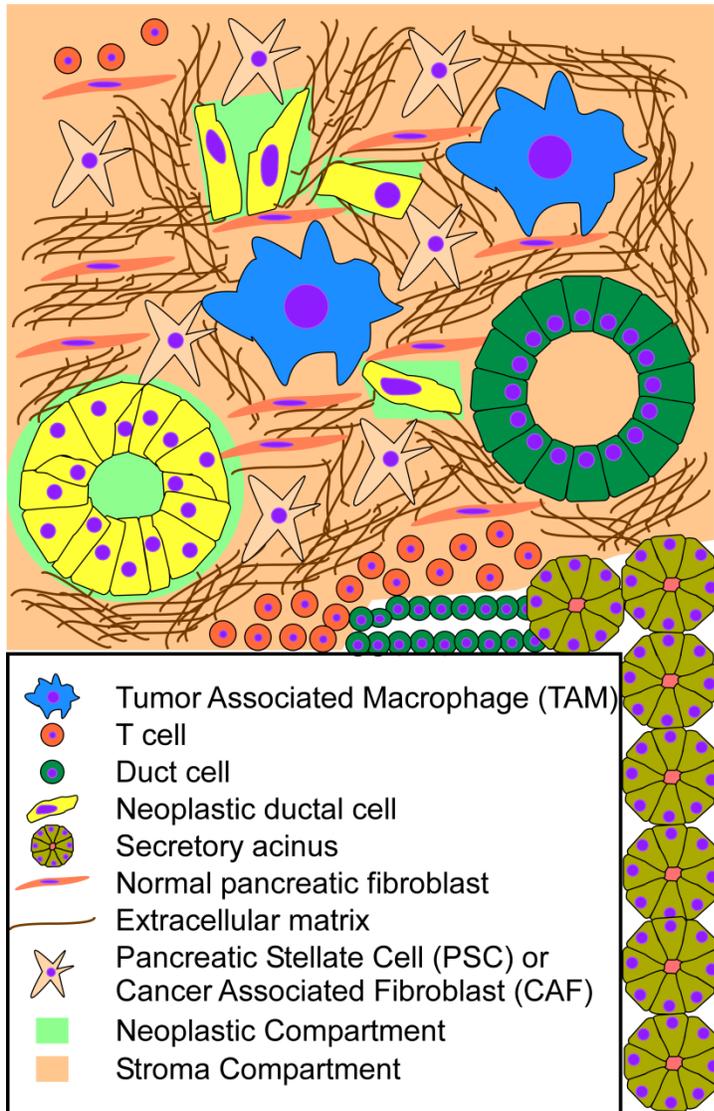
Table 1.2 Cont'd

CA125 <sup>29</sup>	77%	47%
OVA1 <sup>29</sup>	35-40%	94-99%
Colon Cancer		
<i>Fecal Occult Blood Test (FOBT)</i> <sup>21</sup>	98.8%	7.2%
<i>Fecal Immunochemical Test (FIT)</i> <sup>21,22</sup>	87.6-97%	23.2-68%
<i>Cologuard</i> <sup>22</sup>	89.8%	92.3%
<i>Epi procolon</i> <sup>25,26</sup>	81-90%	70-73%
Pancreatic Cancer		
CA19-9 <sup>99</sup>	79%	82%

# In current and former smokers

Chapter 1 Figures

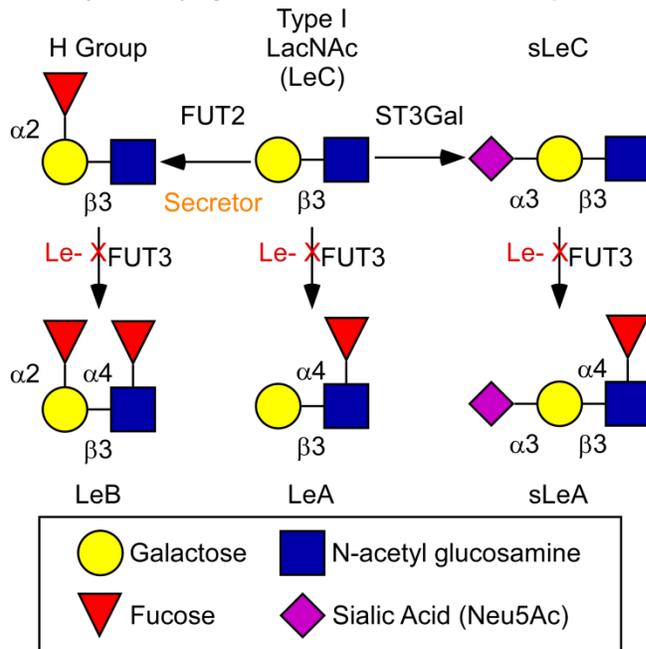
Figure 1.1 Model pancreatic cancer microenvironment.



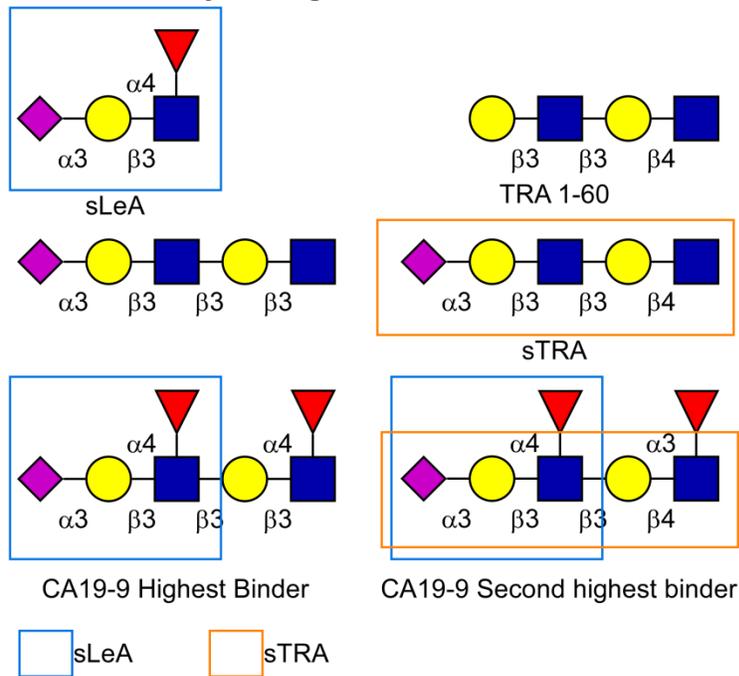
Several major cell types present in most pancreatic tumors are represented in this model. Cells are represented in two significant conceptualized compartments: neoplastic cells and stroma. The vast majority of the pancreatic tumor is stroma comprised of extracellular matrix, activated and quiescent (normal) fibroblasts, and immune cells (tumor associated macrophages, T cells, and other immune cells). Tumor associated macrophages (TAMs) are believed to contribute to active tumor growth and suppression of other immune cells. Activated fibroblasts are also thought to actively signal and contribute to metabolism beneficial to neoplastic cells. Outside of the primary tumor, normal ductal and acinar cells are represented.

Figure 1.2 Glycan synthesis pathway for lewis antigens and likely binding partners for CA19-9.

**A sLeA (CA19-9) Synthesis and Secretor Competition**



**B CA19-9 Antibody Binding**



A. Glycan synthesis pathway for sialyl Lewis A (sLeA), Lewis B (LeB), H group (core blood group antigen, O type), and sTRA is represented with the requisite enzymes STGAL3, FUT2, and FUT3. Lewis secretors have higher expression of FUT2 leading to production of H group over sLeA. Loss of FUT3 results in inability to produce sLeA. B. CA19-9 antibodies bind type I lactosamine containing glycans. They have higher affinity for longer glycans containing the terminal sLeA tetrasaccharide, including the difucosylated variant of sTRA.

## Chapter 2 Tables

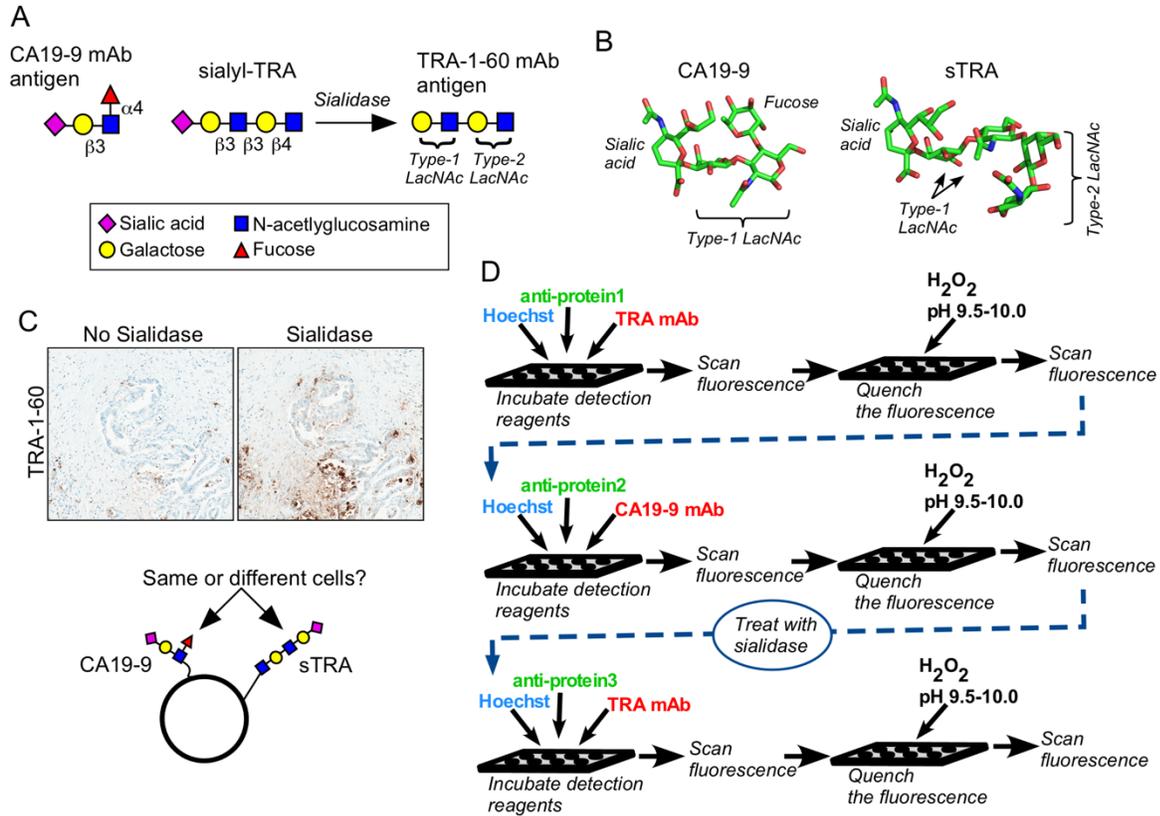
Table 2.1 Results from 10-fold cross validation

	Split1	Split2	Split3	Split4	Split5	Split6	Split7	Split8	Split9	Split10	Median
Round 1	100%	84%	73%	37%	75%	70%	50%	75%	62%	23%	71%
Round 2	53%	88%	42%	80%	53%	100%	55%	88%	88%	80%	80%
Round 3	83%	49%	62%	71%	50%	100%	53%	100%	75%	100%	73%
Round 4	100%	60%	48%	80%	75%	80%	88%	100%	50%	93%	80%
Round 5	80%	96%	100%	35%	71%	68%	88%	100%	92%	70%	84%
Random 1	-	-	-	-	-	-	-	-	-	-	-
Random 2	80%	-	-	-	-	-	-	-	-	-	-
Random 3	-	-	-	-	-	-	-	-	-	-	-
Random 4	-	-	-	-	-	-	-	-	-	-	-
Random 5	-	-	-	-	-	-	-	-	-	-	-

We used the sTRA-only, CA19-9-only, and dual-labeled tissue markers to seek panels that distinguish 30 short-TTP samples from 15 long-TTP samples (Fig. 2.6). Each value is the average accuracy of the panels from the training set applied to the set-aside test samples for each split (see Methods for details). The median accuracy is given for five rounds of cross validation, and the average median was 78%. For the rounds marked 'Random', we randomly assigned a case or control status to each of the 45 samples and repeated the process. Except for Split1 of the second round, the program did not find any panels that met the minimum performance in the training sets.

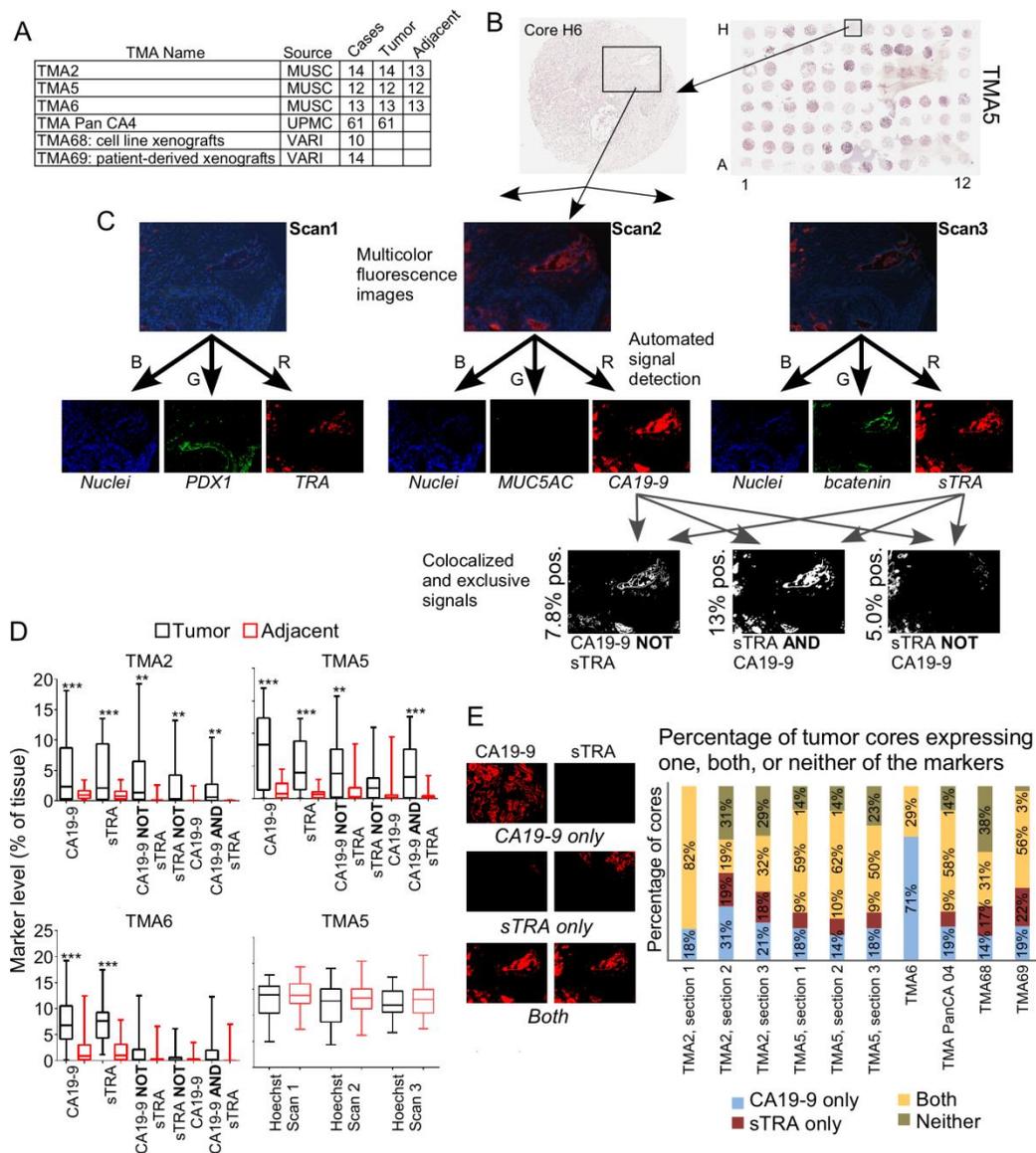
Chapter 2 Figures

Figure 2.1 The CA19-9 and sialyl-TRA (sTRA) antigens.



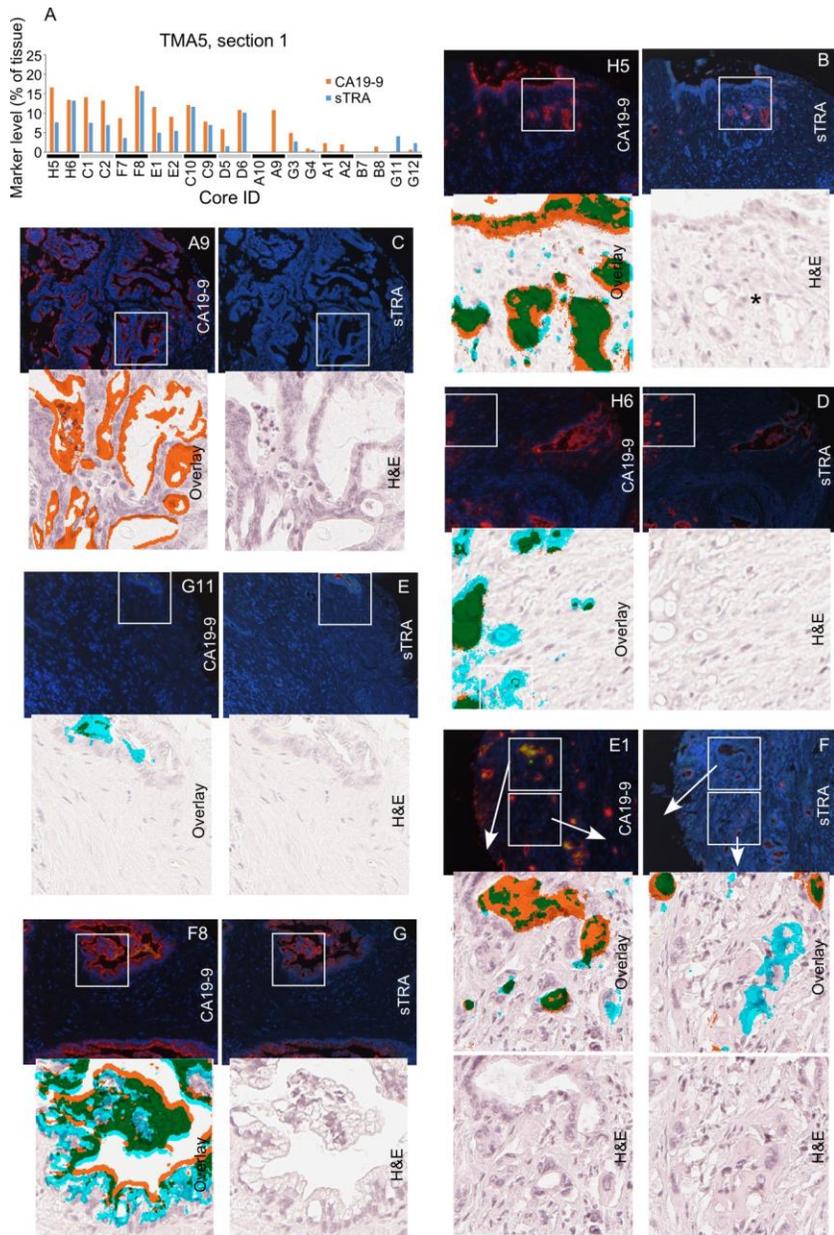
A) In order to detect sTRA, we treat the sample with sialidase prior to applying the TRA-1-60 monoclonal antibody (mAb). B) The CA19-9 and sTRA glycans have similar structures except for the presence of fucose in the CA19-9 antigen. C) The treatment of pancreatic cancer tissue with sialidase leads to increased binding of the TRA-1-60 mAb, revealing the presence of sTRA in the tissue. The overlap with the CA19-9 antigen is not known. D) The schematic shows the process we used for multimarker immunofluorescence. Between the second and third rounds, we treated the tissue with sialidase to remove sialic acid from sTRA, enabling detection by the TRA-1-60 mAb.

Figure 2.2 Quantifying signals in tissue microarrays.



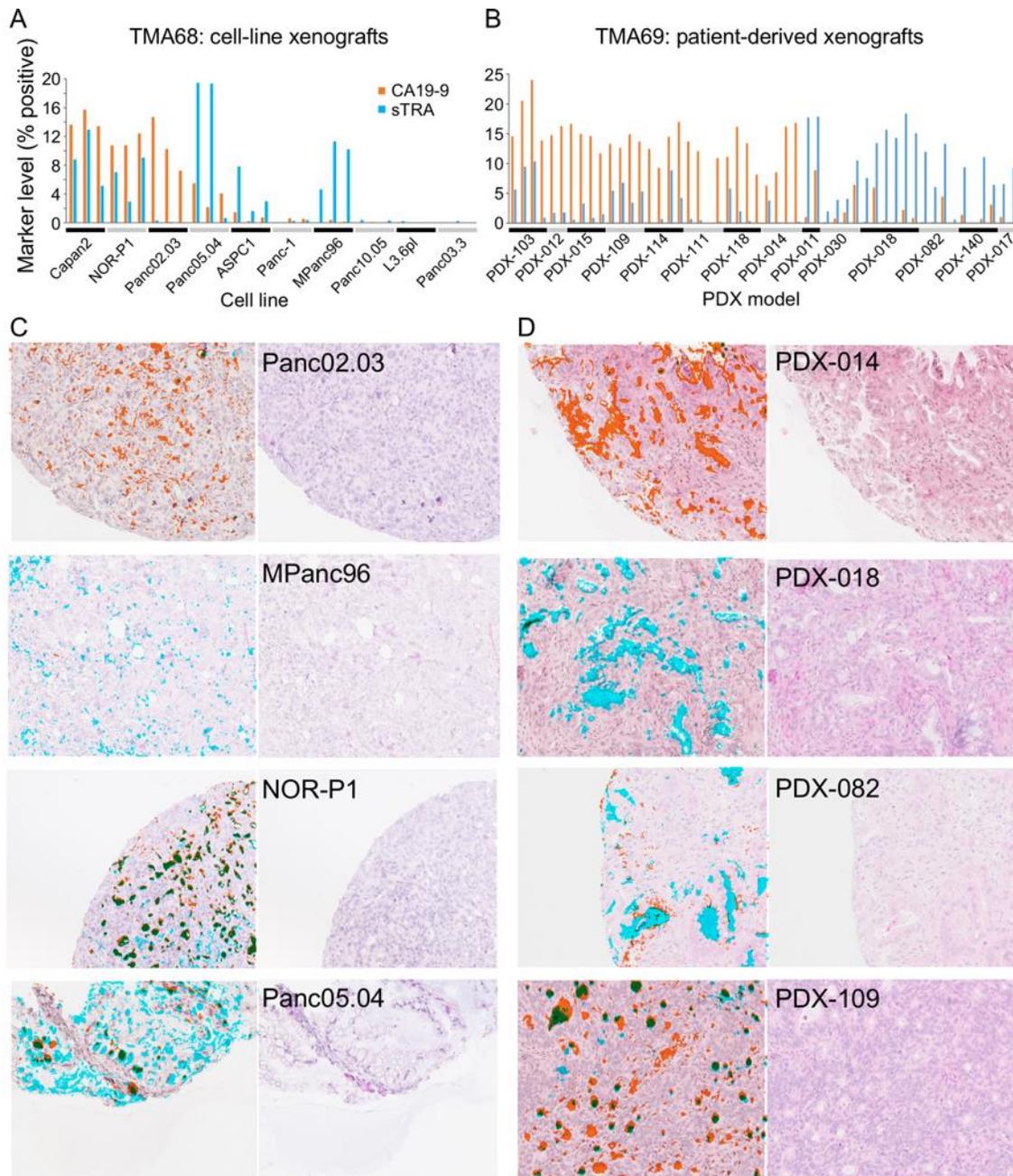
A) We used six TMAs from three sources. Four TMAs contained primary tumor and adjacent tissue, and two contained tumors from xenografts. B) We collected images that were tiled across the TMA. Each image captured a portion of a core, as represented by the box. C) We separately analyzed the blue, green, and red channels from each of the three rounds of immunofluorescence, resulting in nine images per region. The first step was to locate the signals in each image. Next we quantified the amount of signal in each individual image as well as the amount of exclusive or colocalized signal among various combinations of images. D) We compared each of the quantified signals between the tumor cores and the adjacent tissue cores. The amount of signal from the nuclei (stained by Hoechst 33258) was equivalent between tumor and adjacent tissue in all comparisons. E) We quantified the occurrences of tumor tissue that contains one, both, or neither of the markers in each of the TMAs. The images provide examples of each type.

Figure 2.3 Cellular morphologies associated with each glycan.



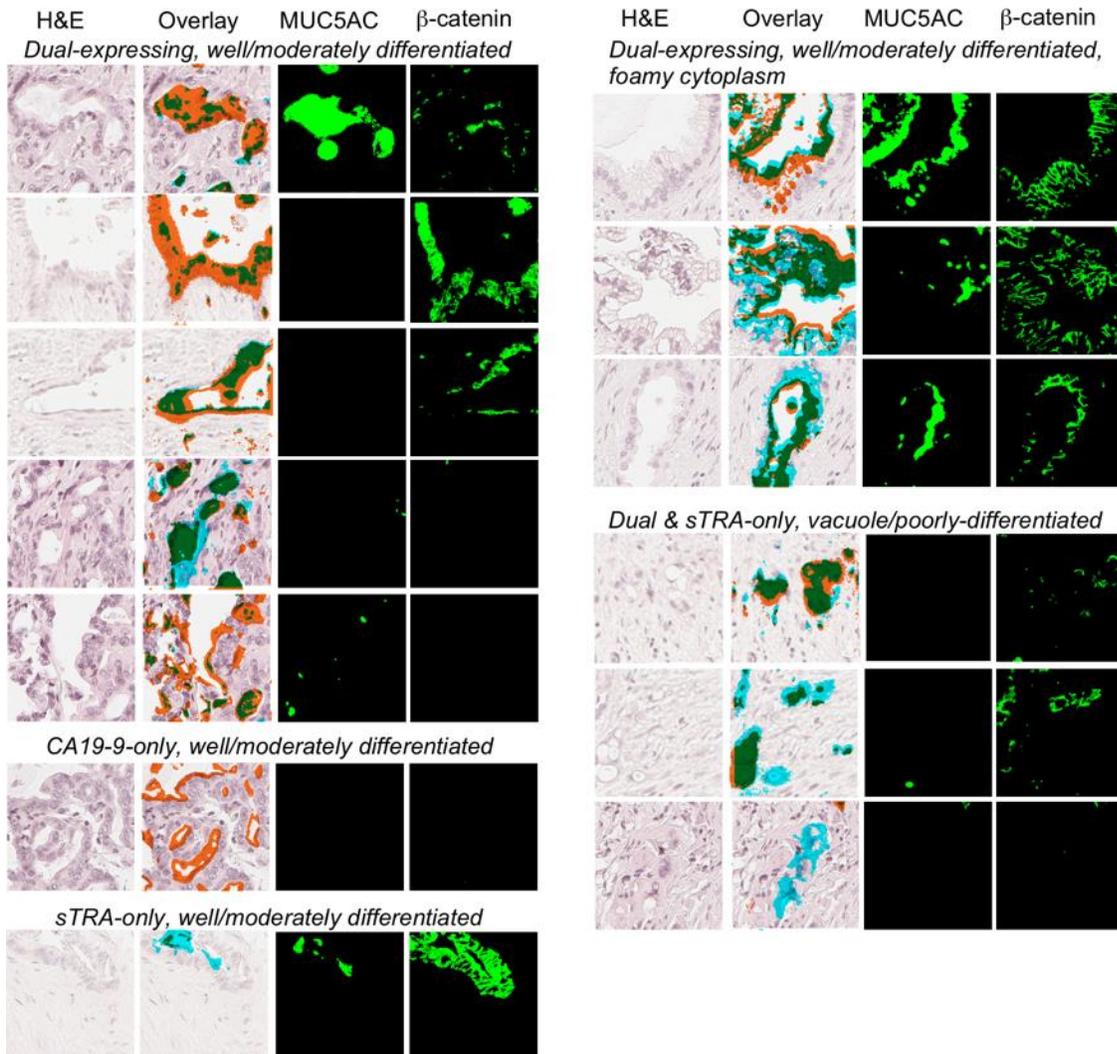
A) The cores had various levels of each antigen. The TMA contained two cores from each tumor, presented in pairs in the column graph. B-G) The top two images in each group are the raw fluorescence from the second (left) and third (right) rounds of immunofluorescence. The red signal in the left image is CA19-9, and the red signal in the right image is sTRA. The lower images in each group are zoomed pictures corresponding to the white box. The lower left image is the H&E image overlaid with the detected signals. The signal from CA19-9 is orange, the signal from sTRA is cyan, and the overlapping signal is green. The lower right image is from the H&E stain. The asterisk in panel B marks vacuolated, invasive cells.

Figure 2.4 Staining and morphologies in xenografts.



A) The cell-line xenograft cores showed various amounts of each antigen. The three cores from each cell line are presented in groups. B) Each PDX xenograft expressed either primarily CA19-9 or primarily sTRA. Each model had 2-6 cores on the TMA. C-D) Selected images from the cell-line (C) and PDX (D) xenografts are presented. In the overlaid images at the right of each pair, CA19-9 is orange, sTRA is cyan, and the overlapping signal is green.

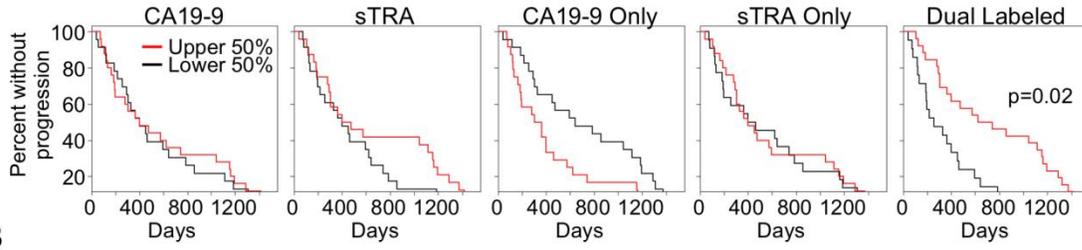
Figure 2.5 Protein expression in various cell types.



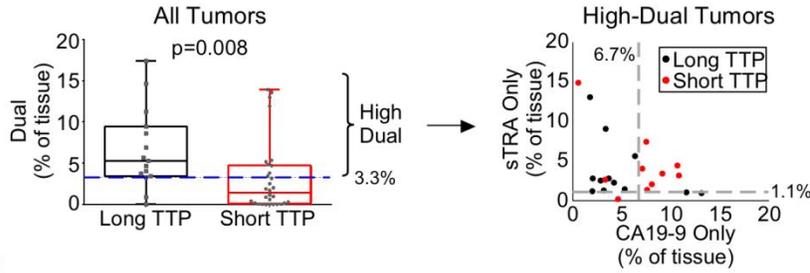
The images are grouped according to glycan expression and morphological phenotype. For each region, we present the H&E image, the H&E with overlaid signal (using the color scheme in figures 3 and 4), the signal detected for MUC5AC, and the signal detected for  $\beta$ -catenin.

Figure 2.6 Associations between glycan type and time-to-progression (TTP).

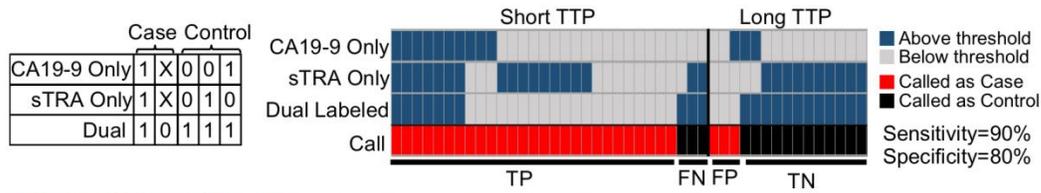
A



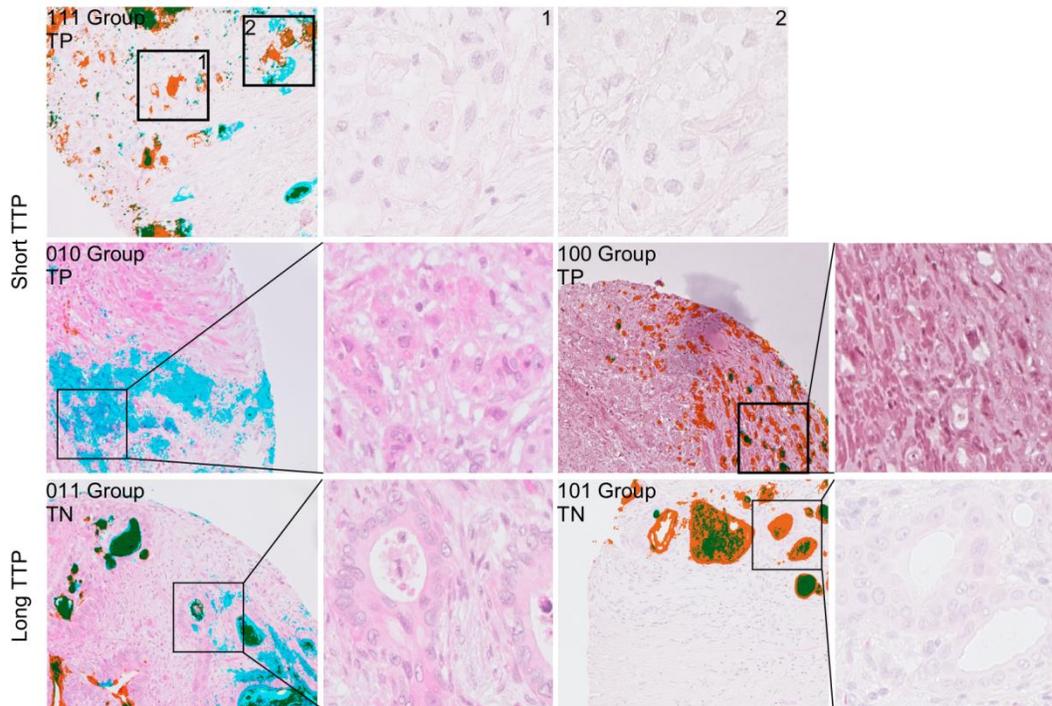
B



C



D



	CA19-9 Only	sTRA Only	Dual
Threshold	6.70	1.11	3.30
111 Average	7.48	7.45	11.92
010 Average	0.15	16.27	0.97
100 Average	10.05	0.92	2.52
011 Average	1.74	13.09	11.27
101 Average	11.58	1.05	5.64

## Figure 2.6 Cont'd

A) We divided the patients into the upper and lower halves of expression of each of the five indicated marker types, and then separately plotted Kaplan-Maier curves for each. B) The tumors with long TTP (>2 years) had significantly higher ( $p = 0.008$ , Wilcoxon rank sum test) dual-marker expression than the patients with short TTP. Among just the tumors with high dual-marker expression, only tumors with short TTP had high levels of both the sTRA-only and CA19-9-only markers. The dashed lines indicate the threshold for each marker, and the value of the threshold is given. C) Patients that were high in all three of the indicated markers, or that were low in the dual marker, were called as cases (short TTP) and all other patients were called as controls (long TTP). The table shows the rules for the calls. A '1' indicates the marker is above its threshold, '0' indicates below threshold, and 'X' indicates either above or below. TP, true positive; FN, false negative; FP, false positive; TN, true negative. D) We present H&E images both with and without overlaid signal. The colors for the overlays are the same as in figures 3-5. A frequent observation among short-TTP patients with high levels of all three markers (top row) or low levels of the dual markers (middle row) was scattered groups of poorly-differentiated cells labeled either with sTRA or CA19-9. A common feature among long-TTP patients (bottom row) was moderately- or well-differentiated PDACs without scattered groups of single-labeled cells. The table provide the averages over all regions for each of the tumors in panel D.

## The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells

**Daniel Barnett,<sup>1\*</sup> Ying Liu,<sup>1\*</sup> Katie Partyka,<sup>1</sup> Ying Huang,<sup>2</sup> Huiyuan Tang,<sup>1</sup> Galen Hostetter,<sup>1</sup> Randall E. Brand,<sup>3</sup> Aatur D. Singhi,<sup>3</sup> Richard R. Drake,<sup>4</sup> and Brian B. Haab<sup>1</sup>**

<sup>1</sup>Van Andel Research Institute, Grand Rapids, MI

<sup>2</sup>Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>3</sup>University of Pittsburgh Medical Center, Pittsburgh, PA

<sup>4</sup>Medical University of South Carolina, Charleston, SC

### Contents

#### Supplementary Tables:

- Table 2A.1. Comparison of glycan expression between tumor and adjacent tissue
- Table 2A.2. Antibody details

#### Supplementary Figures:

- Figure 2A.1. Additional images from the primary tumors
- Figure 2A.2. Additional images from the cell-line xenografts
- Figure 2A.3. Additional images from the PDX xenografts
- Figure 2A.4. E-cadherin and CK19 expression
- Figure 2A.5. Additional images from each tumor group associated with TTP
- Figure 2A.6. Images from tumors with misclassified TTP

## Chapter 2 Supplementary Tables

Table 2A.1 Comparison of biomarker values (averaged over two cores) between tumor and adjacent tissues.

Biomarker	Adjacent Tissue		Tumor Tissue		Difference		Tumor vs. Adjacent	
	Median	IQR	Median	IQR*	Median	IQR	p-value <sup>#</sup>	FDR <sup>\$</sup>
CA19-9	0.438	(0.171,1.006)	5.769	(1.481,11.021)	3.588	(0.574,9.939)	2.22e-08	1.11e-07
CA19-9-only	0.385	(0.085,0.915)	6.089	(1.684,8.069)	4.51	(1.362,7.512)	1.3e-07	3.26e-07
sTRA	0.023	(0.01,0.201)	0.462	(0.026,6.261)	0.246	(-0.005,5.449)	0.00096	0.00096
sTRA-only	0.022	(0.003,0.216)	0.432	(0.021,3.16)	0.3	(0.004,2.547)	0.000764	0.000955
Dual	0.012	(0.001,0.033)	0.762	(0.095,9.546)	0.377	(0.078,10.429)	1.17e-06	1.96e-06

\*Interquantile range

<sup>#</sup>p-value based on the Wilcoxon signed rank test

<sup>\$</sup>False discovery rate

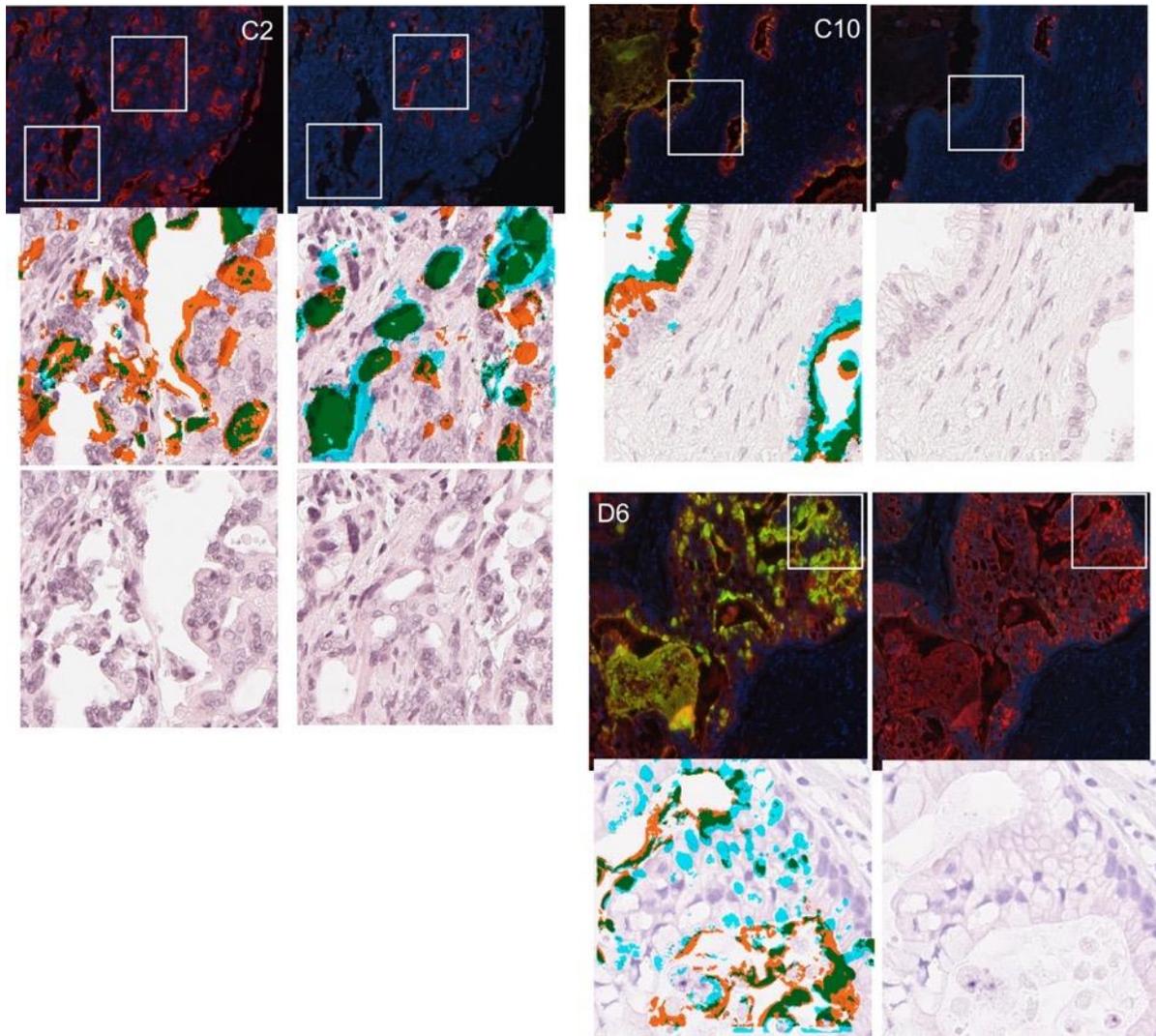
The table presents the median and interquantile range for each biomarker (averaged over two cores) for tumor tissue and adjacent tissue separately, as well as median and interquantile range for their difference in biomarker score. A significant difference between tumor and adjacent tissues was found for each biomarker.

Table 2A.2 Antibody details

Name	Clone ID	Target	Source	Cat. no.	Species	Class	ID
Anti-Sialyl Lewis A (CA19-9)	9L426	Sialyl Lewis A	USBio	C0075-03A	mouse	IgG	1295
TRA-1-60	TRA-1-60	Terminal N-acetyl-lactosamine, type 1	Novus Biologicals	NB100-730	mouse	IgM	1497
Anti-MUC5AC	45M1	MUC5AC	ThermoScientific	MS-145-P1ABX	mouse	IgG1	1480
Anti-beta-catenin	polyclonal	Beta-catenin	R&D Systems	AF1329	goat	IgG	1578
Anti-E-cadherin	3F4	E-cadherin	Sigma Aldrich	WH0000999M1	mouse	IgG1k	1581
Anti-CK19	RCK108	Cytokeratin 19	ThermoScientific	MA1-06329	mouse	IgG1	1591
Anti-vimentin	V9	Vimentin	Sigma Aldrich	V6389	mouse	IgG1	1582
Anti-PDX1	267712	PDX1	R&D Systems	MAB2419	mouse	IgG2B	1551

Chapter 2 Supplementary Figures

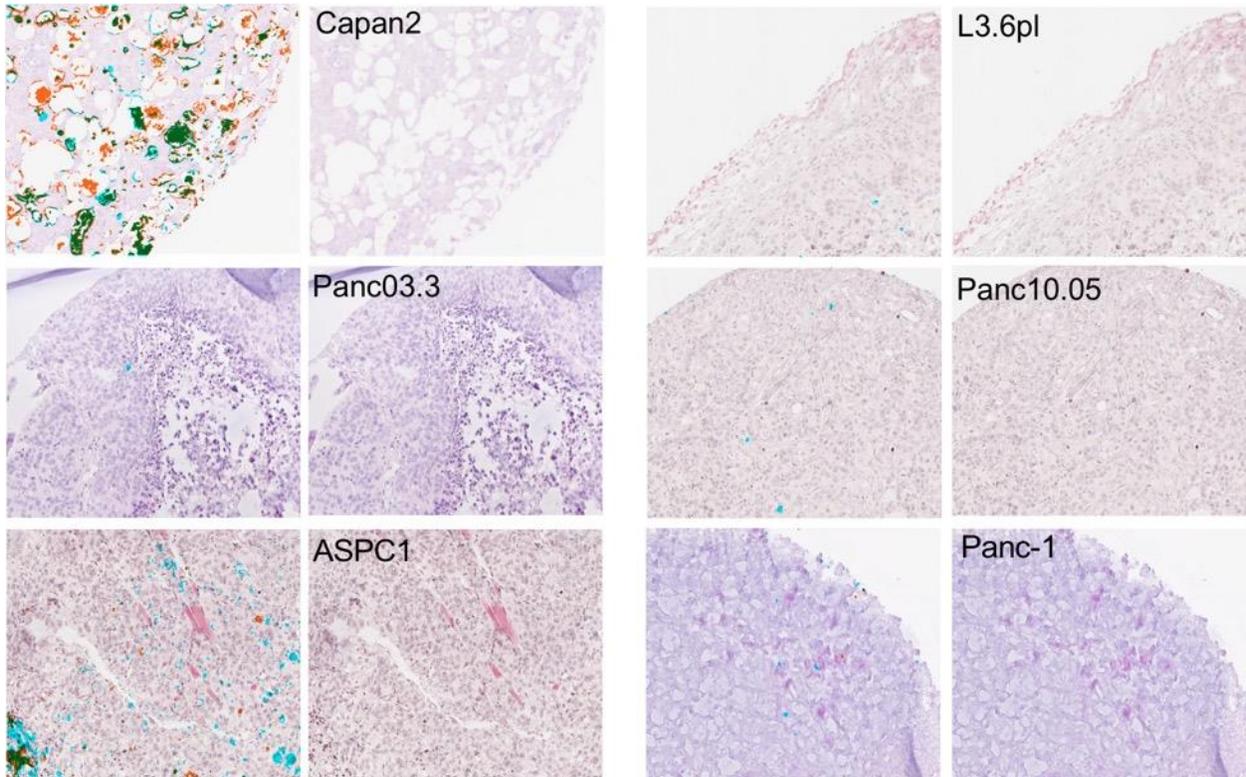
Figure 2A.1 Additional images from the primary tumors



Cores from TMA5 are shown. The multicolor, fluorescence images are at the top of each group, with scan 2 on the left and scan 3 on the right. In scan 2, red was CA19-9 and green was MUC5AC, and in scan 3, red was sTRA and green was  $\beta$ -catenin. Below are H&E images from the area defined by the white box in the fluorescence image. In the overlaid H&Es, orange is CA19-9, cyan is sTRA, and green is the overlap.

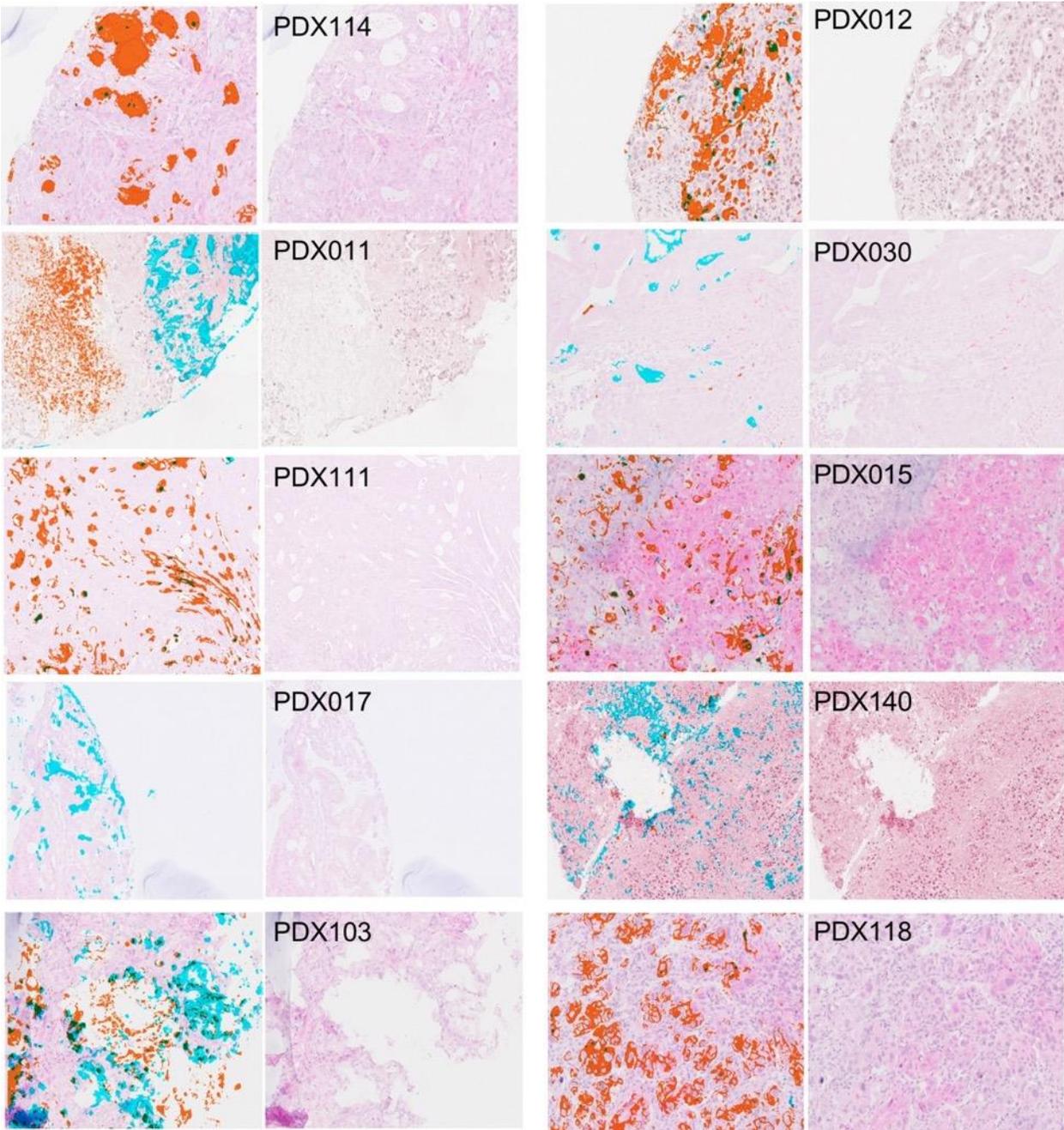
Core C2 shows a moderately-differentiated duct with loose organization that stains mostly with CA19-9 (left), and small glands that secrete dual-labeled material into the lumen (right). Core C10 shows well-differentiated ducts with foamy cytoplasm that generally are labeled with both markers. Core D6 shows lipid-rich and vacuolated cells that label only with sTRA, and clusters that are dual labeled.

Figure 2A.2 Additional images from the cell-line xenografts



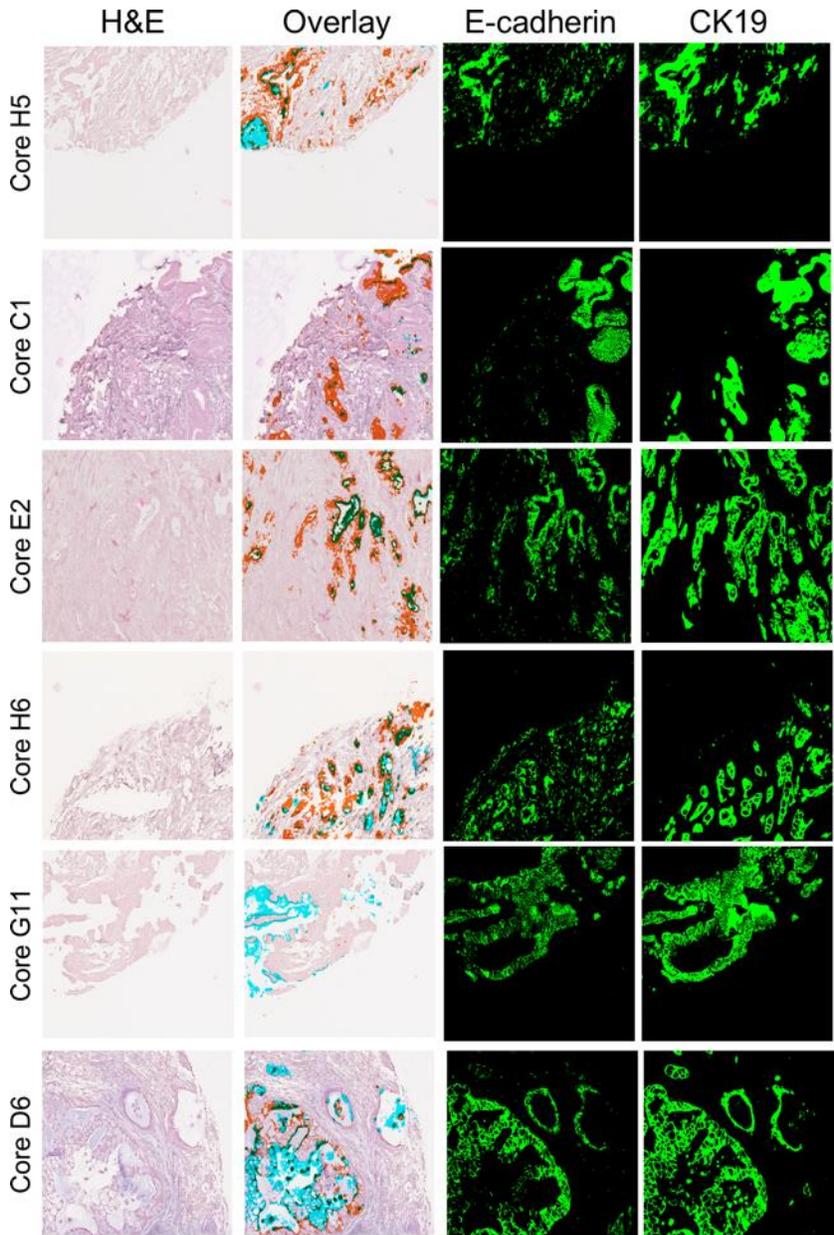
In each pair, the H&E image is on the right and the overlaid image on the left, using the same color scheme as in Figure 2A.1. Capan2 had high staining for both markers, ASPC1 had clear sTRA staining with little CA19-9 staining, and the rest were low in both. The L3.6pl cell line is from adenosquamous carcinoma, not ductal adenocarcinoma like the rest.

Figure 2A.3 Additional images from the PDX xenografts



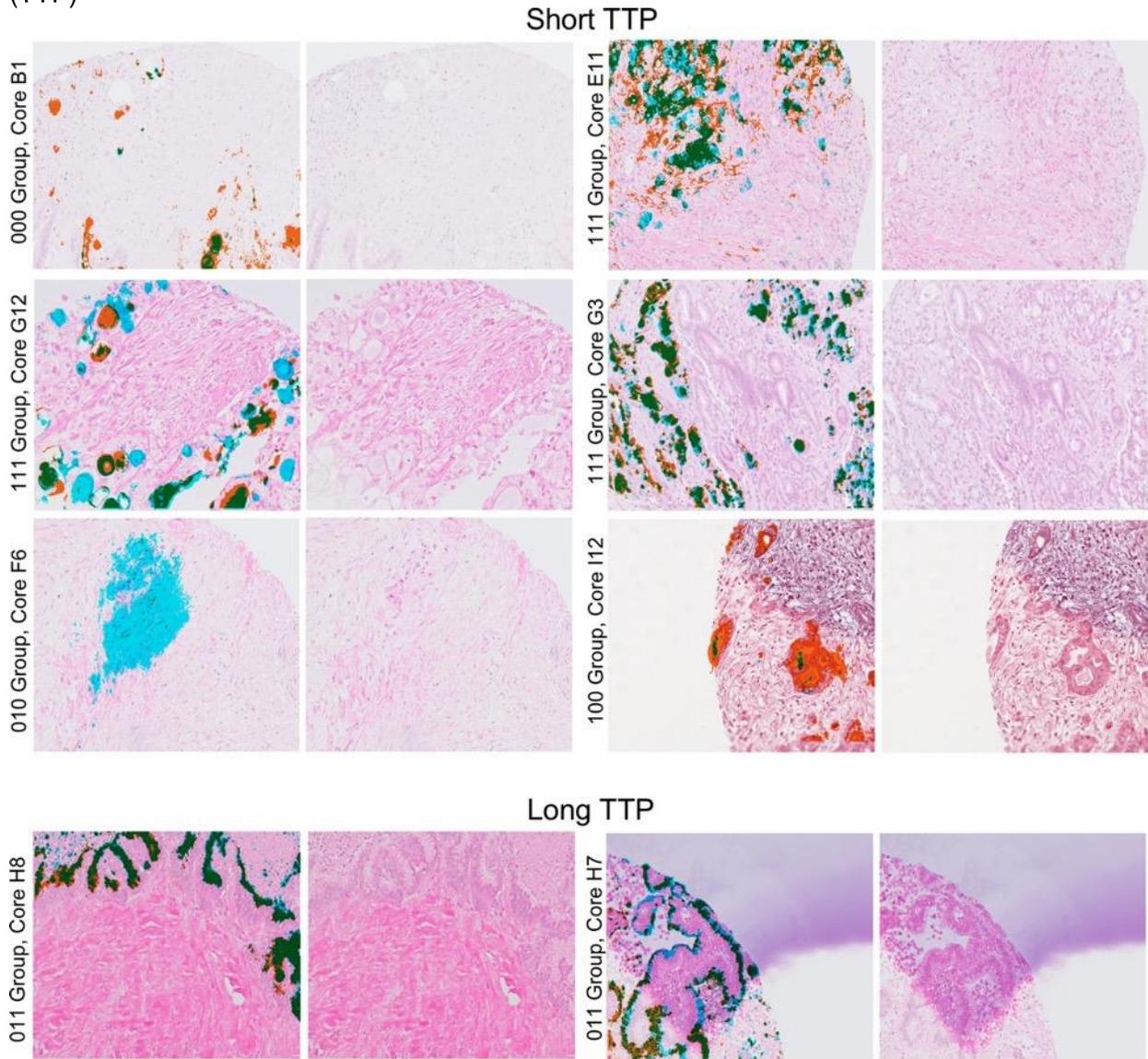
The color scheme of the overlaid images is the same as in the previous figures. Most xenograft models stain primarily with either one or the other marker, and none was wholly absent of staining.

Figure 2A.4 E-cadherin and CK19 expression



The colors for the overlaid H&E are the same as in the previous figures. The right two images show the detected signals corresponding to E-cadherin and CK19. The cells that stained for either sTRA or CA19-9 expressed E-cadherin and CK19, regardless of morphology.

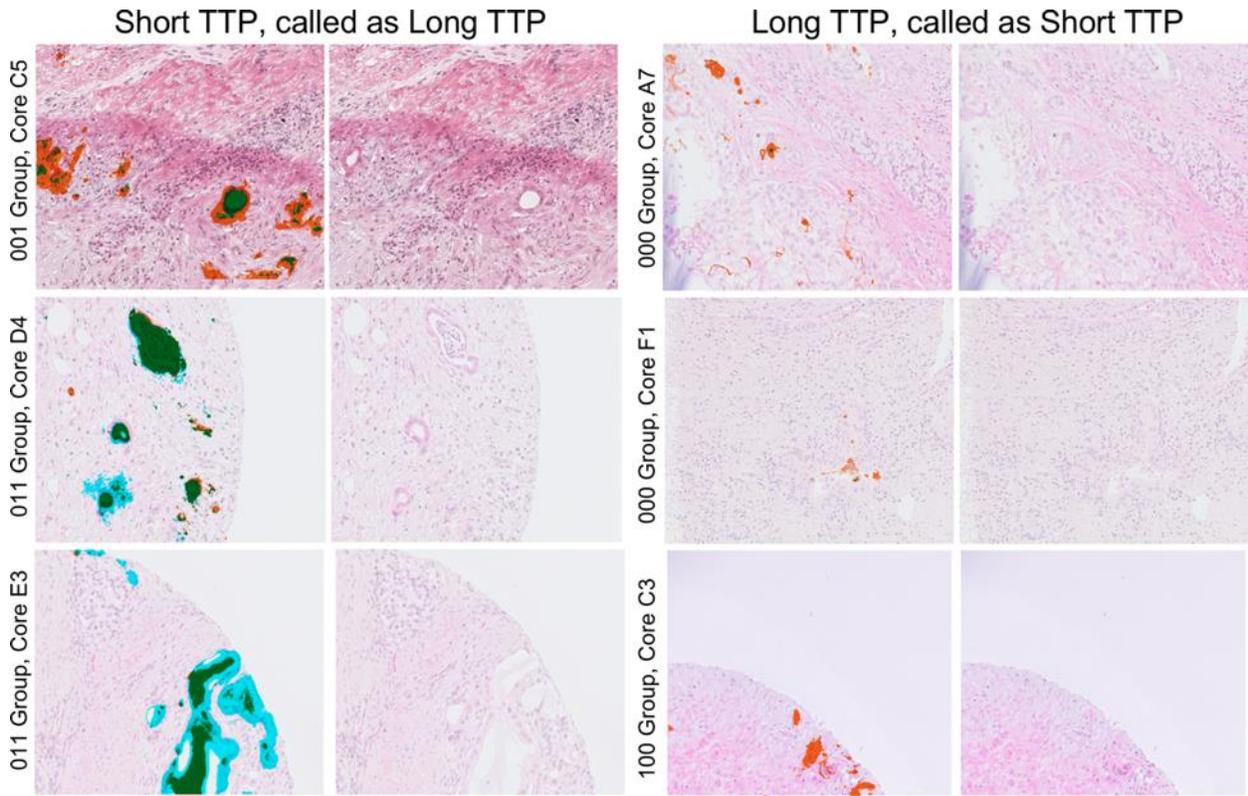
Figure 2A.5 Additional images from each tumor group associated with time-to-progression (TTP)



The color scheme in the overlaid images is the same as in previous figures. The labels on the left give the Core ID and the Group ID according to the table in Figure 2.6C of the main text. The Group ID gives the status of each of the three markers, where a '1' indicates above threshold, and a '0' indicates below threshold. The first number is CA19-9-only, the second is sTRA-only and the third is dual. Thus 000 indicates low in all three markers, 111 indicates high in all three, etc.

The short-TTP tumors are either high in all three markers (cores G12, E11, and G3) or low in the dual-labeled marker (cores B1, F6, and I12). The long-TTP tumors are high in the dual-labeled marker but not in all three of the markers.

Figure 2A.6 Images from tumors with misclassified TTP.



The colors and labeling are the same as in Figure 2A.5.

Table 2A.3 3-marker panel threshold and core averaged marker levels for the images in Figure 2A.5

Marker	CA19-9 Only	sTRA-Only	Dual
Threshold	6.696	1.111	3.303
B1_Region	<b>7.617</b>	0.207	1.529
B1_Patient	5.455	0.115	0.883
E11_Region	12.238	8.778	14.129
E11_Patient	9.128	3.436	5.347
G12_Region	13.058	16.147	10.319
G12_Patient	7.481	7.448	11.921
G3_Region	8.308	8.401	20.564
G3_Patient	10.668	4.453	13.599
F6_Region	0.000	16.399	0.305
F6_Patient	0.001	7.531	0.403
I12_Region	13.422	0.007	0.836
I12_Patient	8.284	0.062	0.594
H8_Region	5.209	7.020	21.905
H8_Patient	3.344	9.079	17.393
H7_Region	6.633	11.922	17.695
H7_Patient	3.344	9.079	17.393

Table 2A.4 3-marker panel threshold and core averaged marker levels for the images in Figure 2A.6

Marker	CA19-9 Only	sTRA-Only	Dual
Threshold	6.696	1.111	3.303
C5_Region	<b>9.133</b>	0.019	3.528
C5_Patient	4.607	0.209	3.823
D4_Region	1.514	3.600	7.350
D4_Patient	3.287	2.649	5.160
E3_Region	0.048	17.277	8.829
E3_Patient	0.539	14.908	4.831
A7_Region	1.762	0.000	0.043
A7_Patient	3.389	0.650	0.876
F1_Region	0.331	0.016	0.018
F1_Patient	0.133	0.035	0.012
C3_Region	<b>4.184</b>	0.000	0.000
C3_Patient	6.859	0.396	3.297

## Chapter 3 Tables

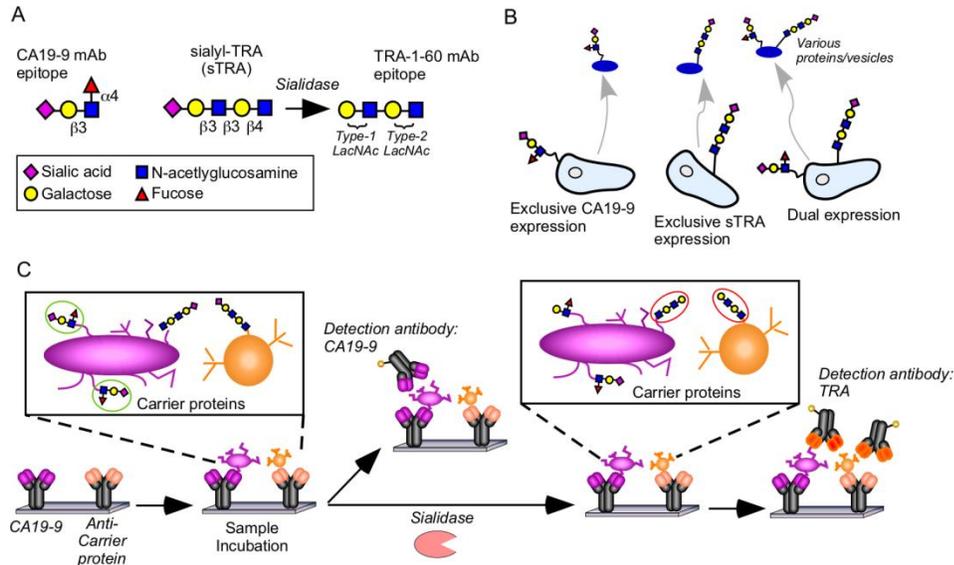
Table 3.1 Composition of the sample sets.

Site	Training/Validation			Test			
	UPMC	UPMC	UPMC	All	UPMC	MDACC	Mayo
Total samples, N	147	50	197 (147 + 50)	147	86	41	20
Cancer, N	72	25	97	71	30	41	0
Average age, y (SD)	65.3 (10.6)	72.8 (8.6)	*67.3 (10.6)	66.3 (9.0)	68.7 (8.6)	64.5 (9.0)	-
Percent male	40	10	51.5%	52.1	50.0	53.7	-
Control, N	75	25	100	76	56	0	20
Average age, y (SD)	57.8 (15.6)	61.8 (15.4)	*58.7 (15.5)	65.0 (10.6)	65.1 (9.2)	-	64 (13.8)
Percent male	34	12	53.0%	44.1	37.5	-	61.9
Cancer stages							
Stage I, N (%)	2 (2.8)	1 (4.0)	3 (3.1)	17 (23.9)	2 (6.7)	15 (36.6)	0
Stage II, N (%)	43 (59.7)	15 (60.0)	58 (59.8)	40 (56.3)	28 (93.3)	12 (29.3)	0
Stage III, N (%)	14 (19.4)	6 (24.0)	20 (20.6)	5 (7.0)	0	5 (12.2)	0
Stage IV, N (%)	13 (18.1)	3 (12.0)	16 (16.5)	9 (12.7)	0	9 (22.0)	0
Control types							
Chronic pancreatitis, N (%)	33 (44.0)	13 (52.0)	46 (46.0)	15 (19.7)	15 (26.8)	0	0
Benign biliary stricture, N (%)	14 (18.7)	9 (36.0)	23 (23.0)	8 (10.5)	8 (14.3)	0	0
Abnormal imaging, N (%)	24 (32.0)	3 (12.0)	27 (27.0)	0	0	0	0
Chronic diabetic, N (%)	0	0	0	24 (31.6)	4 (7.1)	0	20 (100.0)
Healthy control, N (%)	0	0	0	20 (26.3)	20 (35.7)	0	0
Pancreatic cyst, N (%)	4 (5.3)	0	4 (4.0)	9 (11.8)	9 (16.1)	0	0

\*Indicates a significant difference ( $p < 0.001$ , Wilcoxon rank-sum test) between cases and controls. Cells with an em-dash have no value because subjects were not included in that category.

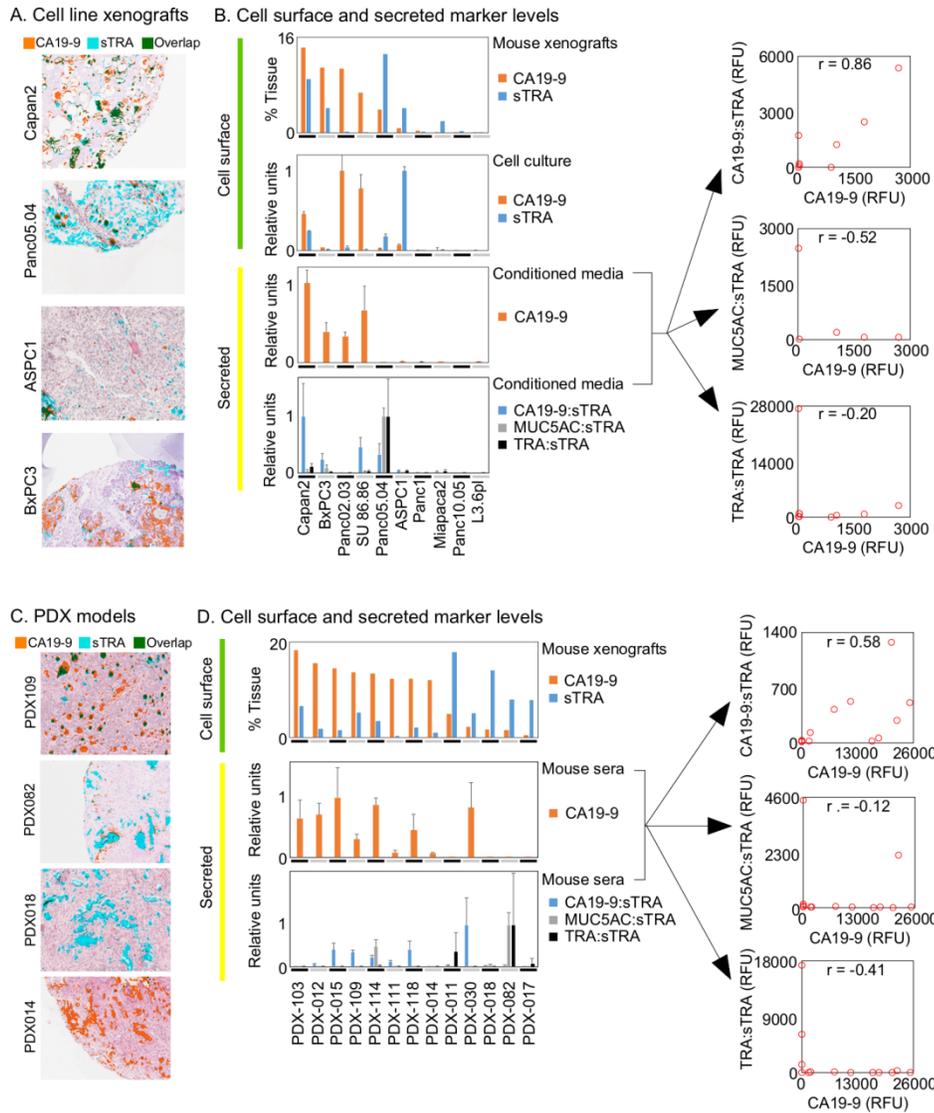
## Chapter 3 Figures

Figure 3.1 The CA19-9 and sTRA assays.



A) The epitopes detected by the CA19-9 and TRA-1-60 antibodies. B) Potential secretion of carriers of single or dual antigens. C) In the CA19-9 assay, both the capture and detection antibodies detect the glycan epitope of the CA19-9 antibody. In the sTRA assay, the capture antibodies target either the CA19-9 antigen or a protein carrier of sTRA. After sample incubation, the captured material is treated with sialidase and then probed with the TRA antibody.

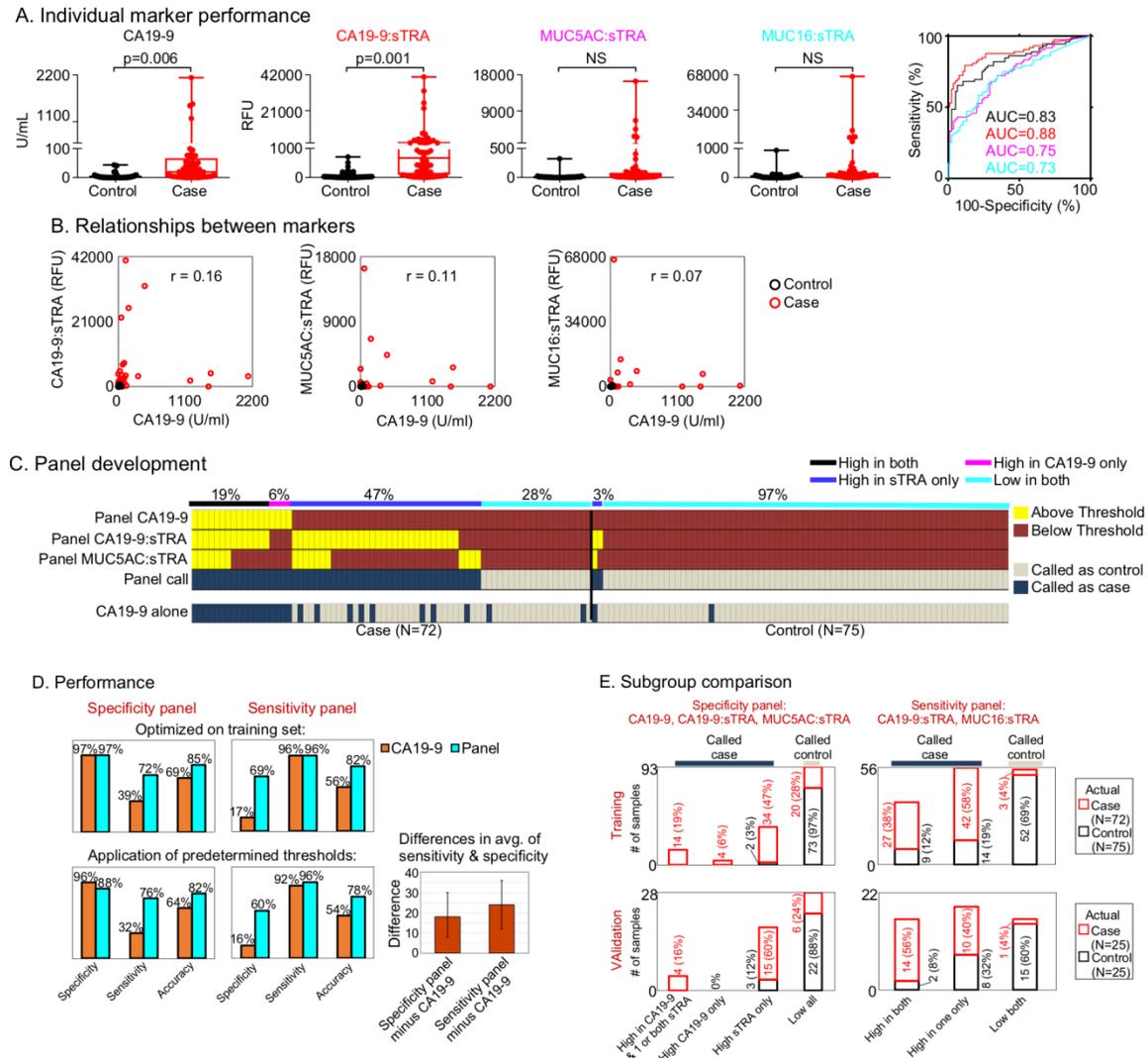
Figure 3.2 Complementary elevations of CA19-9 and sTRA in model systems.



A) Immunofluorescence staining of mouse xenografts of cell lines showed variable expression of the two markers. B) Quantification of the cell surface and secreted levels showed the certain cell lines produced primarily one or the other glycans. C) Immunofluorescence staining of PDX tissue also showed variable expression of the two markers. D) Quantification of the levels in the mouse tissue and sera showed complementary patterns of expression.

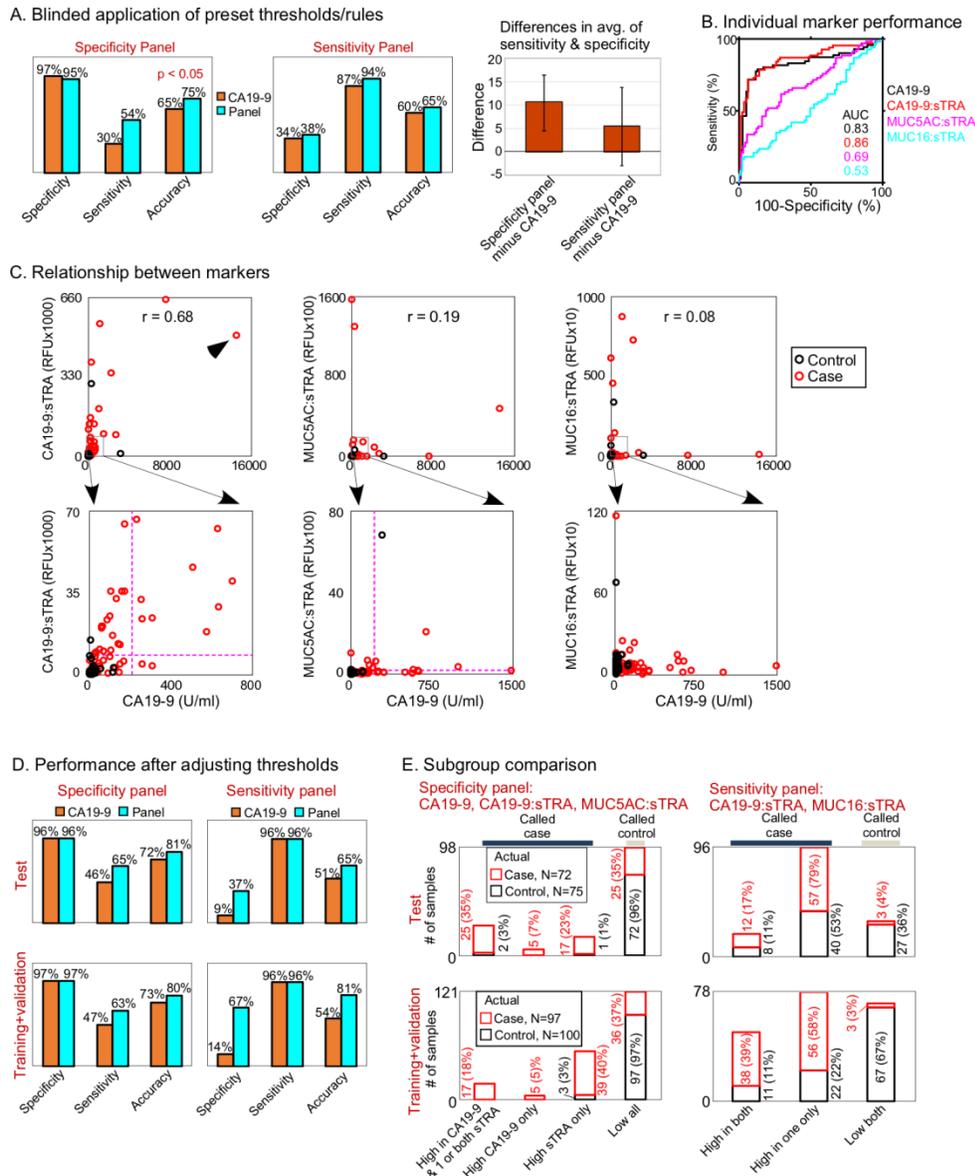


Figure 3.4 Biomarker panel development.



A) The CA19-9 and sTRA assays were quantified in X case and X control plasma specimens. As an single marker, the CA19-9:sTRA assay performed similarly to CA19-9. B) The correlations between the sTRA markers and CA19-9 were very low, with samples elevated in one, both, or neither of the markers. C) A threshold was applied to each marker in the panel or to CA19-9 alone, and samples with an elevation in any marker were called as cases. In the panel optimized for specificity shown here, the panel identified more of the cases than CA19-9. D) The performance of both panels was better than CA19-9 in the training set and in the application of the predetermined thresholds to the 50-sample validation set. For both panels, the difference in the average of sensitivity and specificity was significant ( $p < 0.001$ ). The difference is the average over 1000-fold bootstrapping analysis, and the error bars are the 95% confidence intervals. E) The breakdown of marker contributions and the improvement in final performance were similar between the training and validation sets.

Figure 3.5 Application to blinded samples



The two biomarker panels were applied to a blinded set of 147 samples, using predetermined marker thresholds and classification rules. A) Both panels improved upon CA19-9. The difference in the average of sensitivity and specificity was significant ( $p < 0.001$ ) for the specificity panel, based on 1000-fold bootstrapping analysis. B) The individual marker performances matched the training set. C) The sTRA and CA19-9 markers showed complementary elevations. The higher correlation (0.68) was caused by a sample that was very high in both (arrowhead). The dashed lines show the predetermined thresholds for the specificity panel. D) The improvements in either sensitivity or specificity were consistent between the training and test sets. E) The independent contributions of each panel member and the improvements of the panels over CA19-9 were consistent between the training and test sets.

## Chapter 3 Supplementary Information

“The sTRA Plasma Biomarker: Blinded Validation of Improved Accuracy over CA19-9 in Pancreatic Cancer Diagnosis”

Ben Staal,1\* Ying Liu,1\* Daniel Barnett,1,2\* Peter Hsueh,1 Zonglin He,3 ChongFeng Gao,1 Katie Partyka,1 Mark W. Hurd,6 Aatur D. Singhi,4 Richard R. Drake,5 Ying Huang,3 Anirban Maitra6, Randall E. Brand,4 Brian B. Haab1

1The Van Andel Research Institute, Grand Rapids, MI

2Michigan State University, East Lansing, MI

3Fred Hutchinson Cancer Research Center, Seattle, WA

4University of Pittsburgh Medical Center, Pittsburgh, PA

5Medical University of South Carolina, Charleston, SC

6MD Anderson Cancer Center, Houston, TX

\*Equal contributions

### **Contents**

#### **Supplementary Methods**

Sandwich immunoassays on microarrays

Processing the samples and the biomarker data

Immunofluorescence on tissue microarrays

Cell culture and measuring cell-surface sTRA and CA19-9

Patient-derived xenografts

#### **Supplementary Tables**

Table 3A.1. Training and validation set data.

Table 3A.2. Covariates.

Table 3A.3. Application of the specificity-optimized panel to the test set.

Table 3A.4. Application of the specificity-optimized panel to the test set.

#### **Supplementary Figures**

Figure 3A.1. Correlations between secreted levels and cellular expression.

Figure 3A.2. Individual marker performance in stage I-II and stage III-IV cancers.

### 3A.1 Supplementary Methods

#### 3A.1.1 Sandwich immunoassays on microarrays

We printed forty-eight identical arrays onto glass microscope slides coated with ultra-thin nitrocellulose (PATH Slides, Grace BioLabs). We print microarrays using a contact printer (Aushon 2470, Aushon BioSystems) equipped with 110  $\mu\text{m}$  diameter pins that deposit about 0.3 nL per spot. Each array contained six replicate spots of each antibody in randomized positions within the array. The printed antibodies were CA19-9 (1116-NS-19-9, MyBioSource), anti-MUC5AC (45M1, Thermo Scientific), and anti-MUC16 (X325, Abcam). After printing, hydrophobic borders were imprinted onto the slides (SlidImprinter, The Gel Company, San Francisco, CA) to segregate the arrays and allow for individual sample incubations on each array. The arrays were blocked using 1% bovine serum albumin (BSA) in 1X phosphate buffered saline (1X PBS) plus 0.5% Tween-20 for one hour at room temperature. The slides were rinsed in 1X PBS plus 0.5% Tween-20, washed in the same buffer for 15 minutes, and dried by brief centrifugation at 160 x g, with printed arrays facing outside.

To prepare the plasma samples, they were diluted two-fold or 25-fold into 1X PBS with final concentrations of 0.05% Tween-20, 0.05% Brij-35, an IgG blocking cocktail (100  $\mu\text{g}/\text{mL}$  mouse and rabbit IgG and 50  $\mu\text{g}/\text{mL}$  goat and sheep IgG (Jackson ImmunoResearch)) and protease inhibitor (Complete Mini EDTA-free Tablet, Roche Applied Science). We applied 6  $\mu\text{L}$  of each plasma sample to each array and let the samples incubate overnight at 4 °C. Each unique sample was applied to three separate arrays. The arrays were washed in three changes of PBS/0.1% Tween-20 for three minutes each and dried by centrifugation (Eppendorf 5810R, rotor A-4-62, 1500 x g for three minutes). The arrays to be detected for the sTRA glycan were then treated with  $\alpha$ 2-3 neuraminidase (P0728L, New England Biolabs, Ipswich, MA) at 250 U/mL in the supplied reaction buffer overnight at 37° C.

The following day, the arrays were washed in three changes of PBS/0.1% Tween-20 for three minutes each and dried by centrifugation. We then incubated each array with a biotinylated

detection antibody, prepared at 3 µg/mL in PBS with 0.1% BSA and 0.1% Tween-20. The antibody was either CA19-9 (clone 1116-NS-19-9, MyBioSource) or TRA-1-60 (TRA-160, Novus Biologicals). The biotinylation was performed using a conjugation reagent (EZ-Link Sulfo-NHSBiotin, Thermo Fisher) according to the manufacturer guidelines. After washing and drying the arrays as above, Cy5-conjugated streptavidin (Roche Applied Science) prepared at 2 µg/mL in PBS with 0.1% BSA and 0.1% Tween-20 was incubated for one hour at room temperature, followed by a final wash and dry. We scanned the slides for fluorescence using 633 nm excitation (Innopsys InnoScan 1100 AL).

To quantify the signals, we used in-house software called SignalFinder (available upon request) to locate pixels containing signal in each spot. The program uses the SFT algorithm<sup>129</sup> without user intervention or adjustment of settings. We used a custom script to remove any outliers from the six replicate spots according to the Grubbs' test. The script performs the Grubbs' test for the spot with the greatest deviation from the mean and rejects the spot if the Grubbs' statistic has  $p \geq 0.1$ . The script repeats until either no outliers or only four spots remain and outputs the geometric mean of the non-excluded replicate spots for each array. The script then averages values between replicate arrays.

### *3A.1.2 Processing the samples and the biomarker data*

The samples were run in batches of 50-100 samples. Every sample was run in three replicates in each run, and every sample was run in at least three independent experiments. After quantifying the fluorescence signal for each replicate, the first step was to calibrate the signal. For CA19-9, we used a standard curve that was run in triplicate with each experiment. The fluorescence data from each replicate was calibrated to Units based on the standard curve, and then the value in Units was multiplied by the dilution factor to arrive at a final Units/mL value. The dilution factor was the amount by which the plasma sample was diluted prior to incubation on the array.

To calibrate the sTRA assays, we used a set of 15 calibrator samples. These samples were chosen to cover a range of high, medium, and low values for each of the assays. We used such samples because we had not developed standard material that could be used to produce a calibration curve. The mean fluorescence signal across the calibrator samples was acquired in initial experiments, and this value was taken as the baseline to which future experiments would be calibrated. For each experiment, the mean value of the calibrator samples was determined, and a correction factor was calculated by dividing the baseline mean by the experimental mean. Next, the value for each sample in the experimental set was multiplied by the correction factor to arrive at the final value.

Each experimental batch included a common set of 15 control samples, by which we could assess reproducibility. With each experiment, we determined the correlation across the control samples between the new and previous data, and we determined the CV between replicates in the experiment and between separate experiments.

The next step was to determine whether a sample value was above or below the threshold used in the biomarker evaluations. If the 95% confidence interval of the nine measurements crossed any of the thresholds (from the different panels) for a marker, we ran the assay again for that marker. If the confidence interval crossed the threshold but only one of the technical replicates crossed the threshold with respect to the mean, the sample was not re-run.

To investigate the significance of the improvement, we performed bootstrapping analysis, in which the classification rules are applied to a sampling of the cases and controls over 1000 iterations, and the 95% confidence interval of the difference in performance is determined. All statistical calculations were carried out using the R program, version R-3.2.2

(<https://cran.rproject.org/>).

### *3A.1.3 Threshold adjustment from 1:2 to 1:25 dilutions*

Plasma samples used in sandwich immunoassays are typically diluted into a buffer prior to incubation, and the amount of dilution is determined by the starting concentration of the analyte:

high-concentration analytes require high dilutions to achieve measureable responses, and the opposite is true for low-concentration analytes.

In our previous biomarker research and in the training set, we used 2-fold dilutions to favor detection of the low-concentration analytes. In subsequent optimizations, however, we achieved better reproducibility and more accurate calibration of values using a 25-fold dilution for all assays. Therefore we ran the blinded test set at a 25-fold dilution. The predetermined thresholds derived from the training data could not simply be multiplied by 12.5 (i.e. 25/2), because the changes in marker values were not linear with dilution.

To convert the thresholds derived from the 197-sample training + validation sets to 25-fold dilution values, we ran a subset of the training set in parallel at a 2-fold dilution and at a 25-fold dilution. We then found the threshold in the 25-fold dilution data that provided the same classification—the same samples classified as high or low—as in the 2-fold dilution data. This process was done for each marker, and the resulting predetermined thresholds were applied to the blinded test-set data.

Thus, in this study we used four thresholds for each marker: 1) derived from the 147-sample training set, which were applied to the 50-sample validation set; 2) derived from the combined training + validation sets (197 samples) at the 2-fold dilution; 3) adjusted from the 2-fold to the 25-fold dilution, which were applied to the 147-sample, blinded test set; 4) optimized from the test set. The values are:

		Training (147 samples)	Training + validation (197 samples)	Test (adjusted to 25-fold dilution)	Test (optimized)
Specificity panel	CA19-9	63.10	63.10	1500.00	136.14
	CA19-9:sTRA	398.11	794.33	9929.00	8649.68
	MUC5AC:sTRA	63.10	50.12	220.00	428.55
Sensitivity panel	CA19-9:sTRA	39.81	39.81	470.00	433.51
	MUC16:sTRA	63.10	63.10	100.00	116.68
Specificity	CA19-9 alone	31.06	21.00	208.00	115.52
Sensitivity	CA19-9 alone	0.048	0.022	6.00	0.89

#### 3A.1.4 Immunofluorescence on tissue microarrays

The multimarker-immunofluorescence methods followed those presented earlier<sup>74,159,166</sup>. The tumor specimens were collected from extra portions of surgical resections for pancreatic cancer, and the tissue microarrays were generated from 1 mm cores of formalin-fixed, paraffin embedded (FFPE) samples.

We performed immunofluorescence and chemical stains on 5 µm thick FFPE sections. We removed paraffin by three citrosol washes followed by ethanol/H<sub>2</sub>O rehydration (twice each at 100%, 95%, 70%) and two washes in 1X PBS. We performed antigen retrieval by incubating the slides in citrate buffer at 100°C for 20 minutes, and blocked the slides in 1X phosphate-buffered saline containing 0.05% Tween-20 (PBST0.05) and 3% bovine serum albumin (BSA) for 1 hour at RT. Each round of immunofluorescence was incubated in PBST0.05 with 3% BSA containing two different antibodies (10µg/mL each) (see Table 3A.3 for details about the antibodies), one each labeled with sulfo-Cyanine5 (13320, Lumiprobe) and sulfo-Cyanine3 (11320, Lumiprobe) according to the supplier protocol. We incubated the antibody solution on a tissue section overnight at 4 °C in a humidified chamber.

Next, we decanted the antibody solution and washed the slide three times for 3 minutes each, twice in PBST0.05% and once in 1X PBS. The slide was blotted dry and incubated with Hoechst 33258 (1:1000 dilution in 1X PBS) for 10 minutes at RT. We washed the slides in 1X PBS twice for five minutes and added a coverslip and scanned the slide using a scanning-fluorescence microscope (Vectra, PerkinElmer). The microscope collected 19 images at each field-of-view, each image at a different emission wavelength.

We stored the slides in a humidified chamber until removing the coverslip by slide immersion in deionized water at 37 °C for 30-60 minutes. We quenched the fluorescence by incubating the slide in 6% H<sub>2</sub>O<sub>2</sub> in 250 mM sodium bicarbonate (pH 9.5-10) twice for 20 min. each at RT. The subsequent incubations and scanning steps were as described above.

To treat the slide with sialidase, we incubated a 1:200 dilution (from a 50,000 U/mL stock) of the enzyme ( $\alpha$ 2-3,6,8 Neuraminidase, P0720L, New England Biolabs) in 1X enzyme buffer (5 mM CaCl<sub>2</sub>, 50 mM pH 5.5 sodium acetate) overnight at 37 °C. We washed the slides as above prior to the following antibody incubations. The hematoxylin and eosin (H&E) staining followed a standard protocol with 5.5 - 6 minutes hematoxylin incubation and 3 minutes Eosin incubation. We used in-house software called SignalFinder (available upon request) to locate pixels containing signal in each image. The program uses our SFT algorithm<sup>129</sup> without user intervention or adjustment of settings. From the 19 images captured for each region, we selected the three that corresponded to the emission maxima of Hoechst 33258, Cy3, and Cy5. For each image, SignalFinder creates a map of the locations of pixels containing signal and computes the percentage of tissue-containing pixels that have signal. To arrive at a final number for each core, we averaged over all images for a core. We further analyzed and prepared the data using Microsoft Office Excel and GraphPad Pro, and we prepared the figures using Canvas 14 and Canvas Draw (ACD Systems).

### *3A.1.5 Cell culture and measuring cell-surface sTRA and CA19-9*

All cell lines were cultured in RPMI 1640 medium (Invitrogen) supplemented with 5% fetal bovine serum (Invitrogen). For three-dimensional cell culture, cells were trypsinized and washed with Dulbecco's Phosphate-Buffered Saline, and then suspended in culture medium ( $1 \times 10^7$  cells per mL). The cell suspensions were mixed with Matrigel (Corning) in a 1:3 volume ratio and 50  $\mu$ l of the Matrigel cell suspension were loaded into each well. The cells were feed with 50  $\mu$ l culture medium on top of the Matrigel and cultured for 2-3 days prior to collection of the media for biomarker analysis using the antibody array methods described above.

The following method was used to determine the cell-surface expression of the glycans. Cells were seeded into 96 well plate (2000 cells per well), and cultured for 3 days before fixed in 10% formalin for 20 min. After sialidase treatment (as above), the cells were sequentially incubated with biotin-conjugated TRA-1-60 or CA19-9 antibodies, and streptavidin-conjugated HRP. The

Relative Light Units (RUL) were generated with chemiluminescence reagents (Amersham™ ECL™ Western Blotting Detection Reagents, GE Healthcare), and measured with the plate reader Envision2104 (PerkinElmer). The cells were then stained with 20 μM Hoechst 33258 and the Relative Fluorescence Units (RFU) were measured. The RUL level was normalized with their correspondence RFU of Hoechst 33258 staining. The average of two independent experiments with standard error was presented.

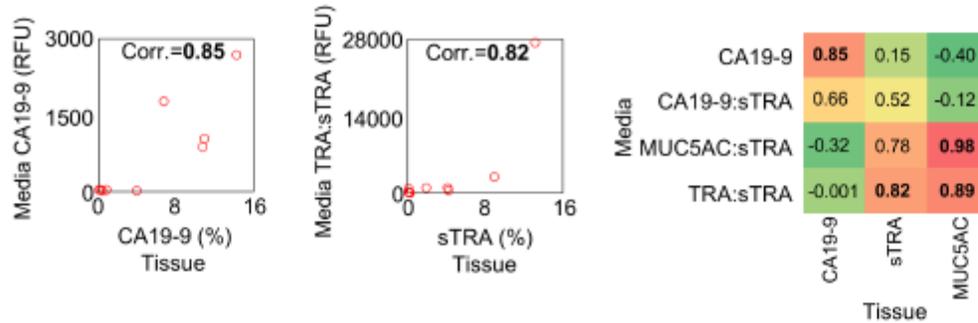
#### *3A.1.6 Patient-derived xenografts*

The tissue and sera from the patient-derived xenografts were from a previously-reported study<sup>155</sup>. The tumor levels of the glycans were determined using immunofluorescence on tissue microarrays, and the biomarker levels in blood serum were determined using antibody sandwich arrays. Both methods are described above.

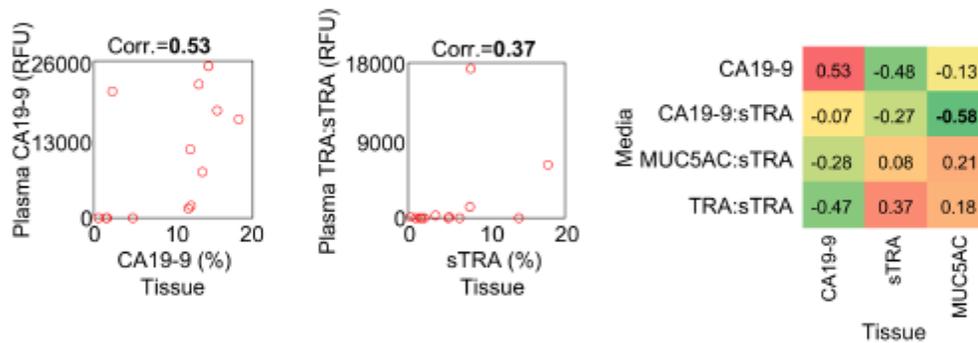
Chapter 3 Supplementary Figures

Figure 3A.1 Correlations between secreted levels and cellular expression.

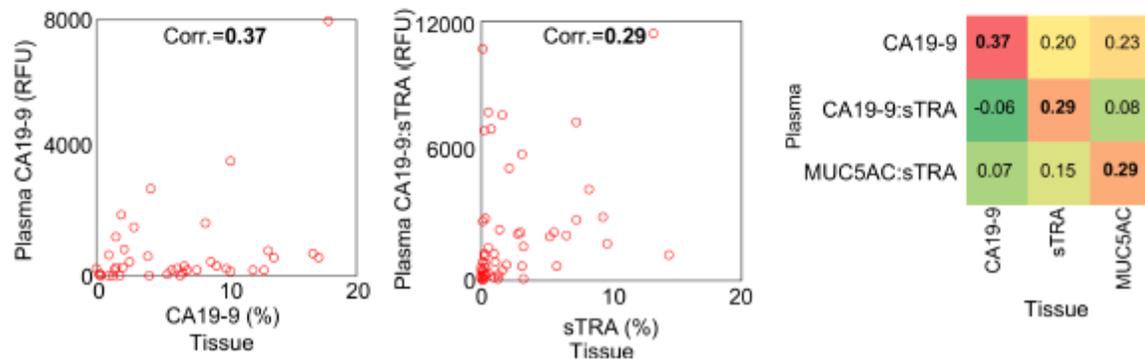
A. Cell lines



B. PDX models



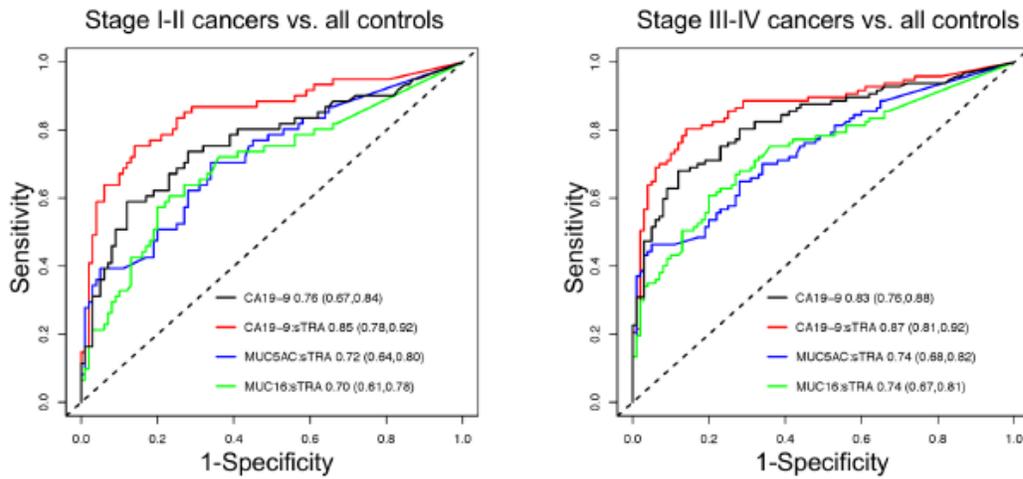
C. Primary specimens



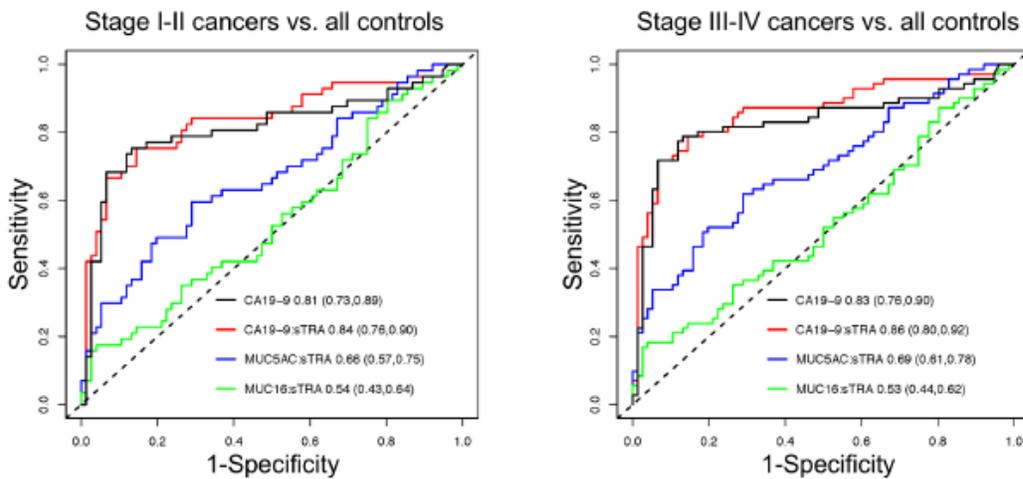
A) The media levels of a marker are plotted with respect to the tissue levels of the corresponding marker for each cell line. Each point on the graph is a unique cell line. The matrix at right shows that the secreted levels of a particular marker correlated with the tissue levels of its corresponding marker, but not with other markers. B) For the PDX models, the serum levels are plotted with respect to the tumor levels. Each point is a unique PDX model. C) For the primary specimens, the blood plasma levels are plotted with respect to the tumor levels, and each point is an individual patient. The three sample types agree in the general correlation of the secreted and tissue levels of a given marker.

Figure 3A.2 Individual marker performance in stage I-II and stage III-IV cancers.

A. Training set



B. Test set



C. AUC values

	Training		Test	
	Stage I-II	Stage III-IV	Stage I-II	Stage III-IV
CA19-9	0.76	0.83	0.81	0.83
CA19-9:sTRA	0.85	0.87	0.84	0.86
MUC5AC:sTRA	0.72	0.74	0.66	0.69
MUC16:sTRA	0.70	0.74	0.54	0.53

The ROC curves for each marker are separately plotted for stage I-II and stage III-IV cancers in the A) training set and the B) test set. C) The summary of area-under-the-curve (AUC) values shows that CA19-9:sTRA was the best-performing marker in each analysis. In addition, the AUCs of all markers were slightly higher in stage III-IV cancers, but not substantially higher.

Chapter 3 Supplementary Tables

Table 3A.1 Training set data

ID	CA19-9	CA19-9:sTRA	MUC5A C:sTRA	MUC16:sTRA	Status (0=control, 1=case)	Diagnosis type	Cancer stage
5144	0.48	116.03	0.00	0.00	0	14 benign stricture; biliary dilation	0
5146	4.81	88.01	13.86	120.04	0	14 benign stricture; biliary dilation	0
5149	0.56	169.97	8.22	7.70	0	10 acute pancreatitis	0
5150	0.40	10.29	0.00	0.00	0	3 chronic pancreatitis	0
5151	0.15	37.40	0.00	0.00	0	10 acute pancreatitis	0
5152	0.58	0.00	0.00	0.00	0	14 benign stricture; biliary dilation	0
5153	0.01	23.56	0.00	5.22	0	14 benign stricture; biliary dilation	0
5156	0.10	0.00	14.95	0.00	0	10 acute pancreatitis	0
5158	0.00	16.64	7.79	10.52	0	3 chronic pancreatitis	0
5159	4.69	35.55	13.59	1.73	0	20 abnormal imaging test (benign)	0
5160	0.00	14.97	5.00	8.13	0	20 abnormal imaging test (benign)	0
5162	3.41	70.58	14.95	0.00	0	11 common bile duct stones	0
5163	2.77	0.00	0.00	12.19	0	11 common bile duct stones	0
5164	2.52	205.72	14.29	101.81	0	5 intraductal papillary mucinous neoplasm (surgical)	0
5165	0.24	0.00	0.00	0.00	0	20 abnormal imaging test (benign)	0
5168	0.23	23.56	0.00	0.00	0	3 chronic pancreatitis	0
5177	0.21	23.56	14.95	0.00	0	20 abnormal imaging test (benign)	0
5178	1.99	42.43	9.23	56.28	0	20 abnormal imaging test (benign)	0
5184	0.52	5.65	5.13	97.10	0	20 abnormal imaging test (benign)	0
5185	41.21	518.51	324.72	939.66	0	3 chronic pancreatitis	0

Table 3A.1 Cont'd

5188	0.00	12.39	2.42	10.27	0	11 common bile duct stones	0
5193	0.41	0.00	0.00	0.00	0	20 abnormal imaging test (benign)	0
5210	16.37	213.77	9.00	49.04	0	3 chronic pancreatitis	0
5215	6.60	3.53	5.97	4.28	0	10 acute pancreatitis	0
5222	0.78	28.73	4.15	7.65	0	20 abnormal imaging test (benign)	0
5224	1.02	87.35	6.02	86.63	0	3 chronic pancreatitis	0
5225	3.80	76.81	0.00	281.05	0	3 chronic pancreatitis	0
5227	0.00	0.00	0.00	9.49	0	3 chronic pancreatitis	0
5229	2.16	49.59	13.48	79.29	0	3 chronic pancreatitis	0
5232	2.76	15.36	3.56	58.53	0	10 acute pancreatitis	0
5233	1.44	23.62	10.77	80.00	0	14 benign stricture; biliary dilation	0
5240	9.02	13.74	2.25	11.79	0	3 chronic pancreatitis	0
5246	2.20	12.11	4.86	0.00	0	3 chronic pancreatitis	0
5248	17.18	73.12	11.33	76.59	0	10 acute pancreatitis	0
5257	0.12	23.56	0.00	0.00	0	11 common bile duct stones	0
5261	43.66	26.53	12.61	59.77	0	10 acute pancreatitis	0
5262	6.47	35.67	3.23	3.99	0	3 chronic pancreatitis	0
5263	5.81	33.17	10.35	39.38	0	20 abnormal imaging test (benign)	0
5264	3.99	52.82	15.27	89.66	0	55 intraductal papillary mucinous neoplasm (clinical)	0
5265	0.64	0.00	0.00	0.00	0	20 abnormal imaging test (benign)	0
5269	3.08	28.10	12.86	65.67	0	3 chronic pancreatitis	0
5278	5.10	0.00	0.00	0.00	0	3 chronic pancreatitis	0
5282	0.03	7.40	12.70	60.96	0	pancreatitis	0

Table 3A.1 Cont'd

5283	0.17	10.23	12.05	61.76	0	14 benign stricture; biliary dilation	0
5292	1.06	87.81	9.40	103.79	0	20 abnormal imaging test (benign)	0
5297	0.04	25.64	15.97	47.31	0	3 chronic pancreatitis	0
5302	0.08	0.00	6.53	5.22	0	10 acute pancreatitis	0
5303	2.13	28.61	15.90	7.16	0	3 chronic pancreatitis	0
5304	0.30	16.48	0.60	5.65	0	20 abnormal imaging test (benign)	0
5308	10.76	19.17	28.29	48.24	0	3 chronic pancreatitis	0
5310	6.17	68.84	15.50	78.12	0	10 acute pancreatitis	0
5312	0.00	28.26	11.89	8.02	0	3 chronic pancreatitis	0
5315	8.97	157.66	10.33	9.31	0	14 benign stricture; biliary dilation	0
5316	0.00	27.07	9.31	37.63	0	10 acute pancreatitis	0
5317	6.33	192.95	0.00	0.00	0	3 chronic pancreatitis	0
5326	0.09	10.95	8.77	10.50	0	14 benign stricture; biliary dilation	0
5329	5.22	27.58	8.70	12.01	0	3 chronic pancreatitis	0
5334	0.62	257.96	0.00	8.55	0	3 chronic pancreatitis	0
5335	9.10	378.59	3.17	3.63	0	16 primary sclerosing cholangitis	0
5337	2.07	88.77	10.09	9.47	0	20 abnormal imaging test (benign)	0
5343	0.25	28.54	5.51	32.42	0	20 abnormal imaging test (benign)	0
5344	0.74	16.05	8.59	51.13	0	20 abnormal imaging test (benign)	0
5358	0.02	13.87	6.91	5.42	0	20 abnormal imaging test (benign)	0
5361	2.21	0.00	0.00	0.00	0	10 acute pancreatitis	0
5364	0.04	19.32	5.96	11.43	0	3 chronic pancreatitis	0

Table 3A.1 Cont'd

5368	0.22	57.51	16.01	31.27	0	55 intraductal papillary mucinous neoplasm (clinical)	0
5369	0.00	79.40	26.82	101.96	0	14 benign stricture; biliary dilation	0
5373	0.45	18.87	14.40	63.11	0	20 abnormal imaging test (benign)	0
5374	20.90	8.77	3.92	4.09	0	20 abnormal imaging test (benign)	0
5376	0.22	9.66	8.23	48.54	0	20 abnormal imaging test (benign)	0
5378	1.40	137.06	17.24	41.84	0	3 chronic pancreatitis	0
5379	0.10	22.02	3.38	5.13	0	20 abnormal imaging test (benign)	0
5382	0.63	17.77	8.56	17.97	0	20 abnormal imaging test (benign)	0
5383	1.25	29.63	10.76	14.04	0	20 abnormal imaging test (benign)	0
5387	19.99	31.93	9.28	6.04	0	55 intraductal papillary mucinous neoplasm (clinical)	0
5389	0.00	10.29	6.93	25.44	0	11 common bile duct stones	0
5393	0.00	0.00	0.00	0.00	0	11 common bile duct stones	0
5394	0.15	27.11	15.83	49.98	0	11 common bile duct stones	0
6069	2.85	0.00	14.95	0.00	0	10 acute pancreatitis	0
6071	1.70	0.00	0.00	0.00	0	10 acute pancreatitis	0
6073	0.56	40.81	0.00	0.00	0	14 benign stricture; biliary dilation	0
6076	0.00	9.62	0.00	0.00	0	10 acute pancreatitis	0
6078	0.78	0.00	4.32	17.24	0	20 abnormal imaging test (benign)	0
6079	0.12	0.00	0.00	0.00	0	20 abnormal imaging test (benign)	0
6084	0.00	23.56	0.00	0.00	0	14 benign stricture; biliary dilation	0
6086	0.00	0.00	0.00	0.00	0	10 acute pancreatitis	0

Table 3A.1 Cont'd

6094	6.19	709.96	19.15	52.04	0	14 benign stricture; biliary dilation	0
6098	0.77	37.40	14.95	13.68	0	10 acute pancreatitis	0
6101	0.84	23.56	0.00	0.00	0	10 acute pancreatitis	0
6102	0.12	23.56	0.00	0.00	0	pancreatitis	0
6105	0.27	7.14	10.57	7.04	0	14 benign stricture; biliary dilation	0
6106	0.62	63.01	0.00	0.00	0	10 acute pancreatitis	0
6134	0.31	0.00	0.00	0.00	0	14 benign stricture; biliary dilation	0
6138	3.30	0.00	0.00	0.00	0	3 chronic pancreatitis	0
6139	1.80	0.00	0.00	0.00	0	20 abnormal imaging test (benign)	0
6142	59.29	3092.98	14.95	0.00	0	14 benign stricture; biliary dilation	0
6144	0.90	403.64	0.00	0.00	0	20 abnormal imaging test (benign)	0
6149	15.04	2783.54	0.00	0.00	0	10 acute pancreatitis	0
6151	2.04	23.56	0.00	17.24	0	3 chronic pancreatitis	0
6157	5.38	23.56	0.00	7.04	0	20 abnormal imaging test (benign)	0
6091	9.56	560.55	8.23	0.00	1	1 pancreatic adenocarcinoma	1
6097	4.04	0.00	0.00	0.00	1	1 pancreatic adenocarcinoma	1
6122	23.80	611.47	16.92	61.46	1	1 pancreatic adenocarcinoma	1
5092	134.35	349.83	13.63	74.61	1	1 pancreatic adenocarcinoma	2
5100	24.59	869.06	12.13	31.69	1	1 pancreatic adenocarcinoma	2
5104	4.00	91.59	14.28	94.81	1	1 pancreatic adenocarcinoma	2
5107	0.61	18.93	14.10	110.31	1	1 pancreatic adenocarcinoma	2
5111	5.31	410.17	4.08	50.66	1	1 pancreatic adenocarcinoma	2
5112	19.67	2236.53	28.76	98.53	1	1 pancreatic adenocarcinoma	2
5113	26.97	4972.93	53.39	118.97	1	1 pancreatic adenocarcinoma	2

Table 3A.1 Cont'd

5114	2.95	9720.25	8383.52	0.00	1	1 pancreatic adenocarcinoma	2
5124	10.43	53.35	17.06	93.68	1	1 pancreatic adenocarcinoma	2
5125	11.98	158.69	0.00	21.72	1	1 pancreatic adenocarcinoma	2
5126	17.10	924.15	1.51	5.15	1	1 pancreatic adenocarcinoma/52 8 pseudopapillary tumor	2
5129	7.28	1976.18	4.28	6.13	1	1 pancreatic adenocarcinoma	2
5130	121.33	3673.97	6.44	11.78	1	1 pancreatic adenocarcinoma	2
5143	0.70	87.29	8.53	3.45	1	1 pancreatic adenocarcinoma	2
5169	6.64	434.34	26.46	54.97	1	1 pancreatic adenocarcinoma	2
5174	5.93	1620.06	9.42	53.09	1	1 pancreatic adenocarcinoma	2
5182	15.22	1339.82	0.00	360.54	1	1 pancreatic adenocarcinoma	2
5191	0.00	863.49	679.14	3483.08	1	1 pancreatic adenocarcinoma	2
5197	19.31	2582.57	43.23	88.92	1	1 pancreatic adenocarcinoma	2
5202	35.89	928.78	0.00	0.00	1	1 pancreatic adenocarcinoma	2
5209	8.33	70.69	0.00	0.00	1	521 intraductal papillary mucinous neoplasm degenerated into adenocarcinoma	2
5219	2.91	0.00	0.00	0.00	1	1 pancreatic adenocarcinoma	2
5221	19.00	1549.39	5.35	6.21	1	1 pancreatic adenocarcinoma	2
5235	16.53	311.48	10.56	49.11	1	1 pancreatic adenocarcinoma	2
5245	7.17	123.51	0.00	0.00	1	1 pancreatic adenocarcinoma	2
5266	1.86	147.21	10.14	8.85	1	1 pancreatic adenocarcinoma	2
5284	0.22	416.20	115.38	145.13	1	1 pancreatic adenocarcinoma/5 intraductal papillary mucinous neoplasm (surgical)	2
5293	39.38	1524.29	10.97	69.00	1	1 pancreatic adenocarcinoma	2
5296	0.02	14.05	13.61	45.91	1	1 pancreatic adenocarcinoma	2

Table 3A.1 Cont'd

5300	7.25	536.47	31.96	84.63	1	1 pancreatic adenocarcinoma	2
5322	51.26	22339.47	154.72	88.86	1	1 pancreatic adenocarcinoma	2
5324	1.95	4133.22	2459.81	137.55	1	1 pancreatic adenocarcinoma	2
5392	0.00	53.56	21.03	589.57	1	1 pancreatic adenocarcinoma	2
6059	45.73	1793.88	34.74	57.69	1	1 pancreatic adenocarcinoma	2
6061	28.67	875.08	2.29	67.08	1	1 pancreatic adenocarcinoma	2
6062	1.55	183.60	11.87	17.24	1	1 pancreatic adenocarcinoma	2
6067	49.86	1147.36	8.23	0.00	1	1 pancreatic adenocarcinoma	2
6074	27.52	534.02	0.00	17.24	1	1 pancreatic adenocarcinoma	2
6081	14.39	1287.64	70.30	17.24	1	1 pancreatic adenocarcinoma	2
6085	3.10	428.03	14.95	0.00	1	1 pancreatic adenocarcinoma	2
6087	0.00	0.00	11.87	0.00	1	1 pancreatic adenocarcinoma	2
6089	67.18	609.54	27.13	57.84	1	1 pancreatic adenocarcinoma	2
6092	101.74	2912.91	404.77	60.33	1	1 pancreatic adenocarcinoma	2
6095	23.46	486.15	10.07	119.98	1	1 pancreatic adenocarcinoma	2
6096	4.38	60.66	18.58	52.59	1	1 pancreatic adenocarcinoma	2
6099	0.39	46.32	318.12	59.30	1	1 pancreatic adenocarcinoma	2
6107	2.30	22.90	23.74	131.40	1	1 pancreatic adenocarcinoma	2
6115	94.83	7677.78	10.94	245.67	1	1 pancreatic adenocarcinoma	2
6117	11.88	453.71	9.51	52.06	1	1 pancreatic adenocarcinoma	2
6121	9.59	99.16	14.95	0.00	1	1 pancreatic adenocarcinoma	2
6128	3.51	25.21	14.25	79.31	1	1 pancreatic adenocarcinoma	2
6143	0.44	189.13	7.82	46.22	1	1 pancreatic adenocarcinoma	2
6145	0.30	39.68	11.69	73.41	1	1 pancreatic adenocarcinoma	2
6146	125.10	40800.00	86.95	7390.89	1	1 pancreatic adenocarcinoma	2
6147	0.08	17.32	10.94	76.59	1	1 pancreatic adenocarcinoma	2

Table 3A.1 Cont'd

6148	40.08	338.59	32.65	75.85	1	1 pancreatic adenocarcinoma	2
6153	82.99	10107.12	59.81	973.95	1	1 pancreatic adenocarcinoma	2
6155	0.04	3918.66	3676.80	8677.99	1	1 pancreatic adenocarcinoma	2
5096	38.58	2155.24	9.10	2.54	1	1 pancreatic adenocarcinoma	3
5116	1.18	7103.85	590.17	118775	1	1 pancreatic adenocarcinoma	3
5122	3.00	10.71	84.84	207.08	1	1 pancreatic adenocarcinoma	3
5135	28.73	1380.86	10.92	102.73	1	1 pancreatic adenocarcinoma	3
5171	29.10	39.90	6.63	5.09	1	1 pancreatic adenocarcinoma	3
5207	132.66	700.71	33.12	173.25	1	1 pancreatic adenocarcinoma	3
5208	25.20	431.13	5.36	73.59	1	1 pancreatic adenocarcinoma	3
5216	49.49	2274.78	23.73	399.84	1	1 pancreatic adenocarcinoma	3
5217	258.69	5900.08	85.68	359.15	1	1 pancreatic adenocarcinoma	3
5223	1.38	0.00	0.00	257.40	1	1 pancreatic adenocarcinoma	3
5226	6.30	1769.80	6323.14	1742.62	1	1 pancreatic adenocarcinoma	3
5236	76.13	7309.85	1.21	7.35	1	1 pancreatic adenocarcinoma	3
5243	69.84	135.08	26.82	31.20	1	1 pancreatic adenocarcinoma	3
5247	99.90	1290.70	0.00	0.00	1	1 pancreatic adenocarcinoma	3
5286	51.77	3815.06	84.03	47.09	1	1 pancreatic adenocarcinoma	3
5299	37.98	141.98	0.00	0.00	1	1 pancreatic adenocarcinoma	3
5306	12.29	870.54	13.26	41.53	1	1 pancreatic adenocarcinoma	3
5355	39.50	2047.09	12.81	54.86	1	1 pancreatic adenocarcinoma	3
5357	166.63	25528.18	6621.20	14193.69	1	1 pancreatic adenocarcinoma	3
5405	1175.5	2045.46	712.99	108.36	1	1 pancreatic adenocarcinoma	3
5099	56.07	15718.97	11234.0	35202.2	1	1 pancreatic adenocarcinoma	4
5103	76.62	390.39	55.64	11.33	1	1 pancreatic adenocarcinoma	4
5105	435.48	32598.42	4419.00	8046.11	1	1 pancreatic adenocarcinoma	4

Table 3A.1 Cont'd

5127	322.25	3496.84	101.05	1263.09	1	1 pancreatic adenocarcinoma	4
5136	58.84	5557.10	46.05	7374.91	1	1 pancreatic adenocarcinoma	4
5139	1476.5	197.54	8.67	47.37	1	1 pancreatic adenocarcinoma	4
5187	313.95	296.49	90.80	0.00	1	1 pancreatic adenocarcinoma	4
5201	10.96	740.82	193.79	202.65	1	1 pancreatic adenocarcinoma	4
5214	82.12	2298.12	5.75	162.45	1	1 pancreatic adenocarcinoma	4
5250	4.20	1565.51	4.86	29.03	1	1 pancreatic adenocarcinoma	4
5259	2.32	9.85	106.08	166.72	1	1 pancreatic adenocarcinoma	4
5285	54.38	241.01	16401.9	66422.7	1	1 pancreatic adenocarcinoma	4
5323	2132.3	3347.95	60.43	517.20	1	1 pancreatic adenocarcinoma	4
5377	19.28	938.69	13.37	84.94	1	1 pancreatic adenocarcinoma	4
5397	27.39	1171.35	3.54	317.21	1	1 pancreatic adenocarcinoma	4
5398	1515.2	4312.47	2684.11	6820.12	1	1 pancreatic adenocarcinoma	4

Table 3A.2 Analysis of covariates between the biomarker data and clinical information

Training set

Age		Case			Control		
		N	corr	p-value	N	corr	p-value
	CA19-9	97	-0.14	0.17	100	0.28	0.0046
	CA19-9:sTRA	97	0.087	0.4	100	0.11	0.29
	MUC5AC:sTRA	97	0.043	0.68	100	0.019	0.85
	MUC16:sTRA	97	-0.007	0.95	100	-0.004	0.97

Test set

Age		Case			Control		
		N	corr	p-value	N	corr	p-value
	CA19-9	71	-0.003	0.98	76	0.12	0.32
	CA19-9:sTRA	71	-0.052	0.67	76	0.14	0.24
	MUC5AC:sTRA	71	-0.1	0.39	76	0.3	0.0077
	MUC16:sTRA	71	0.091	0.45	76	-0.18	0.13

Training set

Gender		Case				Control			
		N(%)	median	IQR	p-value	N (%)	median	IQR	p-value
CA19-9	Female	47(48.5)	8.3	(2.3,33.0)	0.0033	54(54.0)	0.7	(0.2,2.9)	0.69
	Male	50(51.5)	28.9	(9.6,80.6)		46(46.0)	0.8	(0.1,5.1)	
CA19-9: sTRA	Female	47(48.5)	700.7	(135.4,1873.0)	0.52	54(54.0)	23.6	(6.1,42.0)	0.45
	Male	50(51.5)	872.8	(150.1,2830.3)		46(46.0)	23.6	(11.7,67.4)	
MUC5AC: sTRA	Female	47(48.5)	14.9	(8.4,78.0)	0.62	54(54.0)	5.3	(0.0,12.8)	0.89
	Male	50(51.5)	13.9	(7.0,58.7)		46(46.0)	8	(0.0,10.5)	
MUC16: sTRA	Female	47(48.5)	73.6	(14.3,187.9)	0.93	54(54.0)	6.5	(0.0,51.5)	0.67
	Male	50(51.5)	64.3	(18.4,133.2)		46(46.0)	9.5	(0.0,45.9)	

Test set

Gender		Case				Control			
		N(%)	median	IQR	p-value	N (%)	median	IQR	p-value
CA19-9	Female	34(47.9)	102.3	(24.9,520.6)	0.59	43(56.6)	7	(5.6,14.4)	0.49
	Male	37(52.1)	106.1	(33.5,207.7)		33(43.4)	6.8	(5.3,12.3)	
CA19-9: sTRA	Female	34(47.9)	9399.5	(1799.8,57059.8)	0.95	43(56.6)	617.9	(355.9,1214.7)	0.66
	Male	37(52.1)	13217.1	(2781.1,35766.5)		33(43.4)	568.3	(335.6,924.8)	
MUC5AC: sTRA	Female	34(47.9)	129.1	(70.0,190.0)	0.9	43(56.6)	79.2	(57.6,112.6)	0.98
	Male	37(52.1)	113.8	(78.3,183.6)		33(43.4)	80.5	(53.3,114.5)	
MUC16: sTRA	Female	34(47.9)	52.5	(36.9,73.6)	0.27	43(56.6)	57.8	(32.6,74.9)	0.62
	Male	37(52.1)	63.7	(41.2,92.9)		33(43.4)	60.7	(40.7,94.5)	

Table 3A.2 Cont'd

Training set

Stage		N(%)	median	IQR	p-value
CA19-9	I-II	61(62.9)	9.6	(2.3,27.0)	0.000066
	III-IV	36(37.1)	53.1	(23.7,41.2)	
CA19-9: sTRA	I-II	61(62.9)	486.1	(91.6,549.4)	0.021
	III-IV	36(37.1)	1473.2	(366.9,357.6)	
MUC5AC: sTRA	I-II	61(62.9)	13.6	(8.2,32.0)	0.052
	III-IV	36(37.1)	39.6	(8.2,12.8)	
MUC16: sTRA	I-II	61(62.9)	57.7	(8.8,93.7)	0.005
	III-IV	36(37.1)	164.6	(38.9,70.3)	

Test set

Stage		N(%)	median	IQR	p-value
CA19-9	I-II	57(80.3)	90.9	(22.4,173.9)	0.023
	III-IV	14(19.7)	308.2	(69.4,927.4)	
CA19-9: sTRA	I-II	57(80.3)	8901.4	(1743.8,32717.4)	0.017
	III-IV	14(19.7)	32219.4	(11077.4,177285.9)	
MUC5AC: sTRA	I-II	57(80.3)	109.2	(67.4,173.0)	0.045
	III-IV	14(19.7)	156.7	(103.9,1844.7)	
MUC16: sTRA	I-II	57(80.3)	59.2	(38.1,92.0)	0.57
	III-IV	14(19.7)	52.8	(33.5,104.7)	

Table 3A.2 Cont'd

Key

Control types	
1	Healthy
2	Benign stricture
3	Chronic diabetes
4	Chronic pancreatitis
5	Cyst

Training set

Control type		N(%)	median	IQR	p-value
CA19-9	1	21(21.0)	0.8	(0.1,2.2)	0.63
	2	23(23.0)	0.5	(0.0,5.5)	
	3	27(27.0)	0.7	(0.2,1.9)	
	4	25(25.0)	2.2	(0.2,5.2)	
	5	4(4.0)	1.6	(0.5,6.9)	
CA19-9: sTRA	1	21(21.0)	23.6	(0.0,37.4)	0.68
	2	23(23.0)	23.6	(10.3,83.7)	
	3	27(27.0)	18.9	(9.2,32.6)	
	4	25(25.0)	25.6	(15.4,73.1)	
	5	4(4.0)	44.7	(23.9,94.6)	
MUC5AC: sTRA	1	21(21.0)	3.2	(0.0,9.3)	0.61
	2	23(23.0)	6.9	(0.0,13.2)	
	3	27(27.0)	6	(2.0,9.7)	
	4	25(25.0)	7.8	(0.0,12.7)	
	5	4(4.0)	11.8	(7.0,14.7)	
MUC16: sTRA	1	21(21.0)	0	(0.0,7.7)	0.05
	2	23(23.0)	9.3	(0.0,37.7)	
	3	27(27.0)	9.5	(2.9,53.7)	
	4	25(25.0)	12	(7.2,49.0)	
	5	4(4.0)	18.7	(4.5,48.9)	

Test set

Control type		N(%)	median	IQR	p-value
CA19-9	1	20(26.3)	7.6	(6.1,10.1)	0.000026
	2	8(10.5)	16.1	(7.3,22.2)	
	3	24(31.6)	2.4	(0.8,7.9)	
	4	15(19.7)	19.5	(7.8,42.2)	
	5	9(11.8)	6.9	(6.7,8.4)	

Table 3A.2 Cont'd

CA19-9: sTRA	1	20(26.3)	373.8	(249.7,687.2)	0.0061
	2	8(10.5)	1401.5	(383.8,2756.8)	
	3	24(31.6)	610.4	(446.2,973.4)	
	4	15(19.7)	1096	(607.0,2115.3)	
	5	9(11.8)	325.4	(183.5,412.4)	
MUC5AC: sTRA	1	20(26.3)	71.9	(57.1,98.2)	0.9
	2	8(10.5)	87.1	(66.7,111.5)	
	3	24(31.6)	89	(57.8,108.8)	
	4	15(19.7)	91.3	(51.2,106.1)	
	5	9(11.8)	79.2	(57.5,119.4)	
MUC16: sTRA	1	20(26.3)	56.4	(39.3,76.2)	0.078
	2	8(10.5)	63.2	(47.1,78.5)	
	3	24(31.6)	42.7	(28.4,63.3)	
	4	15(19.7)	69.9	(55.5,99.1)	
	5	9(11.8)	65.2	(42.0,91.0)	

p-value methods	
Age	Spearman's correlation
Gender	Wilcoxon rank sum test
Stage/control type	Kruskal-Wallis rank sum test

Pooled cancer stages	
I-II	I, IA, IB, II, IIA, IIB
III-IV	III-IV: III, IIIA, IV

Table 3A.3 Application of the specificity panel and CA19-9 to blinded samples.

ID	Status	Marker values			Preset panel thresholds			Preset CA19-9-alone threshold	Blinded panel markers, thresholded			Blinded panel call	Blinded CA19-9 call	Blinded results*	
		CA19-9	CA19-9: TRA	MUC5AC : TRA	CA19-9	CA19-9: sTRA	MUC5AC : sTRA	CA19-9	CA19-9	CA19-9: sTRA	MUC5AC : sTRA	Call	Call	Panel	CA19-9
6437	1	0.51	469.77	119.01	1500	9929	220	208	0	0	0	0	0	FN	FN
6438	1	14.04	1743.82	56.15	1500	9929	220	208	0	0	0	0	0	FN	FN
6440	1	24.74	1777.63	44.97	1500	9929	220	208	0	0	0	0	0	FN	FN
6441	1	25.29	1530.89	71.59	1500	9929	220	208	0	0	0	0	0	FN	FN
6444	1	207.72	161706.3	16488.63	1500	9929	220	208	0	1	1	1	0	TP	TP
6446	1	151.35	12864.24	75.21	1500	9929	220	208	0	1	0	1	0	TP	FN
6447	1	48.49	3779.54	128.27	1500	9929	220	208	0	0	0	0	0	FN	FN
6451	1	1001.8	199156.9	365.13	1500	9929	220	208	0	1	1	1	1	TP	TP
6453	1	629.34	62686.21	205.91	1500	9929	220	208	0	1	0	1	1	TP	TP
6456	1	57.23	20587.6	88.73	1500	9929	220	208	0	1	0	1	0	TP	FN
6458	1	33.63	111672.9	157443.8	1500	9929	220	208	0	1	1	1	0	TP	FN
6459	1	44.98	4914.95	108.47	1500	9929	220	208	0	0	0	0	0	FN	FN
6460	1	257.66	23806.19	142.3	1500	9929	220	208	0	1	0	1	1	TP	TP
6464	1	309.58	24258.2	113.76	1500	9929	220	208	0	1	0	1	1	TP	TP
6466	1	254.47	32267.11	127.04	1500	9929	220	208	0	1	0	1	1	TP	TP
6469	1	106.09	10481.75	183.62	1500	9929	220	208	0	1	0	1	0	TP	FN
6472	1	704.18	40180.51	2102.54	1500	9929	220	208	0	1	1	1	1	TP	TP
6473	1	635.92	29143.2	135.57	1500	9929	220	208	0	1	0	1	1	TP	TP
6474	1	5.01	471.37	63.05	1500	9929	220	208	0	0	0	0	0	FN	FN
6475	1	26.34	7865.06	175.55	1500	9929	220	208	0	0	0	0	0	FN	FN
6476	1	100.08	25170.23	138.53	1500	9929	220	208	0	1	0	1	0	TP	FN
6478	1	24.14	1111.44	105.72	1500	9929	220	208	0	0	0	0	0	FN	FN
6481	1	58.57	19827.56	84.12	1500	9929	220	208	0	1	0	1	0	TP	FN
6482	1	306.86	3827.78	129.87	1500	9929	220	208	0	0	0	0	1	FN	TP
6484	1	84.78	23539.76	129.25	1500	9929	220	208	0	1	0	1	0	TP	FN
6485	1	0.44	1866.3	69.48	1500	9929	220	208	0	0	0	0	0	FN	FN
6489	1	7561.44	652598.6	100.65	1500	9929	220	208	1	1	0	1	1	TP	TP
6490	1	33.47	9651.47	123.7	1500	9929	220	208	0	0	0	0	0	FN	FN
6491	1	508.03	45991.96	127.73	1500	9929	220	208	0	1	0	1	1	TP	TP

Table 3A.3 Cont'd

6492	1	22.43	8901.39	182.72	1500	9929	220	208	0	0	0	0	0	FN	FN
6494	1	167.3	5771.59	79.36	1500	9929	220	208	0	0	0	0	0	FN	FN
6496	1	2206.45	346997.5	9368.55	1500	9929	220	208	1	1	1	1	1	TP	TP
6497	1	14543.3	504125	48031.74	1500	9929	220	208	1	1	1	1	1	TP	TP
6503	1	148.38	4603.99	78.33	1500	9929	220	208	0	0	0	0	0	FN	FN
6505	1	67.52	5936.85	79.82	1500	9929	220	208	0	0	0	0	0	FN	FN
6506	1	1.81	923.41	51.95	1500	9929	220	208	0	0	0	0	0	FN	FN
6507	1	38.43	1071.31	102.38	1500	9929	220	208	0	0	0	0	0	FN	FN
6508	1	1.27	169.98	38.57	1500	9929	220	208	0	0	0	0	0	FN	FN
6510	1	0.44	5932.9	1071.21	1500	9929	220	208	0	0	1	1	0	TP	FN
6515	1	2724.86	89836.24	2730.36	1500	9929	220	208	1	1	1	1	1	TP	TP
6517	1	146.68	964.77	168.04	1500	9929	220	208	0	0	0	0	0	FN	FN
6526	1	10.36	2497.84	105.55	1500	9929	220	208	0	0	0	0	0	FN	FN
6531	1	230.71	66730.45	634.52	1500	9929	220	208	0	1	1	1	1	TP	TP
6540	1	173.95	35766.54	139.32	1500	9929	220	208	0	1	0	1	0	TP	FN
6541	1	141.15	13217.11	96.02	1500	9929	220	208	0	1	0	1	0	TP	FN
6544	1	132.12	134067.7	11864.44	1500	9929	220	208	0	1	1	1	0	TP	FN
6545	1	582.93	137330.3	250.04	1500	9929	220	208	0	1	1	1	1	TP	TP
6547	1	5.97	529.8	54.91	1500	9929	220	208	0	0	0	0	0	FN	FN
6550	1	260.31	4636.38	172.99	1500	9929	220	208	0	0	0	0	1	FN	TP
6551	1	8.18	1067.06	74.85	1500	9929	220	208	0	0	0	0	0	FN	FN
6558	1	104.54	35703.93	42.96	1500	9929	220	208	0	1	0	1	0	TP	FN
6561	1	173.93	64733.53	490.6	1500	9929	220	208	0	1	1	1	0	TP	FN
6562	1	112.04	16640.75	124.05	1500	9929	220	208	0	1	0	1	0	TP	FN
6564	1	1495.94	94374.81	192.47	1500	9929	220	208	0	1	0	1	1	TP	TP
6571	1	212.5	80906.49	317.37	1500	9929	220	208	0	1	1	1	1	TP	TP
6572	1	50.6	10747.94	133.31	1500	9929	220	208	0	1	0	1	0	TP	FN
6576	1	576.79	18532.79	58.66	1500	9929	220	208	0	1	0	1	1	TP	TP
6577	1	16.69	4227.35	66.1	1500	9929	220	208	0	0	0	0	0	FN	FN
6580	1	65.38	9897.59	104.1	1500	9929	220	208	0	0	0	0	0	FN	FN
6583	1	5.62	251.08	67.44	1500	9929	220	208	0	0	0	0	0	FN	FN
6587	1	158.49	35792.78	696.63	1500	9929	220	208	0	1	1	1	0	TP	FN
6597	1	124.96	2781.09	48.72	1500	9929	220	208	0	0	0	0	0	FN	FN
6598	1	62.84	20628.93	99.1	1500	9929	220	208	0	1	0	1	0	TP	FN

Table 3A.3 Cont'd

6598	1	62.84	20628.93	99.1	1500	9929	220	208	0	1	0	1	0	TP	FN
6600	1	99.39	7586.51	46.48	1500	9929	220	208	0	0	0	0	0	FN	FN
6601	1	1106.73	551586.2	14852.55	1500	9929	220	208	0	1	1	1	1	TP	TP
6602	1	333.57	391359.7	130431.8	1500	9929	220	208	0	1	1	1	1	TP	TP
6603	1	129.18	32717.41	62.28	1500	9929	220	208	0	1	0	1	0	TP	FN
6607	1	5.04	430.56	63.02	1500	9929	220	208	0	0	0	0	0	FN	FN
6609	1	7.08	109.36	109.16	1500	9929	220	208	0	0	0	0	0	FN	FN
6612	1	90.9	588.99	66.43	1500	9929	220	208	0	0	0	0	0	FN	FN
6616	1	7.24	398.73	156.2	1500	9929	220	208	0	0	0	0	0	FN	FN
6518	0	8.51	339.49	57.52	1500	9929	220	208	0	0	0	0	0	TN	TN
6520	0	23.8	887.54	84.37	1500	9929	220	208	0	0	0	0	0	TN	TN
6521	0	5.51	482.85	115.21	1500	9929	220	208	0	0	0	0	0	TN	TN
6525	0	8.37	240.04	74.91	1500	9929	220	208	0	0	0	0	0	TN	TN
6533	0	15.32	1637.83	37.27	1500	9929	220	208	0	0	0	0	0	TN	TN
6534	0	5.58	795.71	70.92	1500	9929	220	208	0	0	0	0	0	TN	TN
6535	0	5.74	354.79	127.17	1500	9929	220	208	0	0	0	0	0	TN	TN
6536	0	12.89	651.06	92.49	1500	9929	220	208	0	0	0	0	0	TN	TN
6556	0	6.95	224.03	28.66	1500	9929	220	208	0	0	0	0	0	TN	TN
6559	0	13.35	218.81	82.06	1500	9929	220	208	0	0	0	0	0	TN	TN
6569	0	11.35	357	72.8	1500	9929	220	208	0	0	0	0	0	TN	TN
6570	0	6.83	1028.47	264.99	1500	9929	220	208	0	0	1	1	0	FP	TN
6575	0	6.6	368.93	55.69	1500	9929	220	208	0	0	0	0	0	TN	TN
6579	0	9.7	241.32	140.17	1500	9929	220	208	0	0	0	0	0	TN	TN
6582	0	5.53	421.24	69.42	1500	9929	220	208	0	0	0	0	0	TN	TN
6584	0	8.27	252.54	21.5	1500	9929	220	208	0	0	0	0	0	TN	TN
6588	0	6.21	204.23	66.94	1500	9929	220	208	0	0	0	0	0	TN	TN
6594	0	8.31	6965.61	191.44	1500	9929	220	208	0	0	0	0	0	TN	TN
6608	0	6.16	572.5	59.97	1500	9929	220	208	0	0	0	0	0	TN	TN
6615	0	5.26	378.69	25.74	1500	9929	220	208	0	0	0	0	0	TN	TN
6528	0	27.02	2475.54	100.34	1500	9929	220	208	0	0	0	0	0	TN	TN
6548	0	5.26	120.42	73.92	1500	9929	220	208	0	0	0	0	0	TN	TN
6553	0	20.64	2048.46	65.37	1500	9929	220	208	0	0	0	0	0	TN	TN
6563	0	6.46	201.82	50.67	1500	9929	220	208	0	0	0	0	0	TN	TN
6566	0	15.9	444.4	109.91	1500	9929	220	208	0	0	0	0	0	TN	TN
6591	0	16.26	3600.64	152.96	1500	9929	220	208	0	0	0	0	0	TN	TN

Table 3A.3 Cont'd

6599	0	3181.49	12401.13	116.14	1500	9929	220	208	1	1	0	1	1	FP	FP
6613	0	7.56	754.58	67.19	1500	9929	220	208	0	0	0	0	0	TN	TN
6439	0	0.54	388.13	99.94	1500	9929	220	208	0	0	0	0	0	TN	TN
6443	0	5.21	602.88	53.31	1500	9929	220	208	0	0	0	0	0	TN	TN
6445	0	2.02	617.89	58.04	1500	9929	220	208	0	0	0	0	0	TN	TN
6449	0	1.79	1119.3	118.46	1500	9929	220	208	0	0	0	0	0	TN	TN
6450	0	2.8	442.07	45.3	1500	9929	220	208	0	0	0	0	0	TN	TN
6455	0	4.33	696.99	101.49	1500	9929	220	208	0	0	0	0	0	TN	TN
6462	0	8.43	1722.52	179.65	1500	9929	220	208	0	0	0	0	0	TN	TN
6463	0	1.91	14674.27	76.49	1500	9929	220	208	0	1	0	1	0	FP	TN
6465	0	0.44	300.26	21.9	1500	9929	220	208	0	0	0	0	0	TN	TN
6471	0	0.38	645.39	135.26	1500	9929	220	208	0	0	0	0	0	TN	TN
6483	0	1.25	177.05	102.35	1500	9929	220	208	0	0	0	0	0	TN	TN
6486	0	0.34	388.55	53.07	1500	9929	220	208	0	0	0	0	0	TN	TN
6487	0	2.84	335.6	114.45	1500	9929	220	208	0	0	0	0	0	TN	TN
6493	0	12.33	661.67	154.45	1500	9929	220	208	0	0	0	0	0	TN	TN
6498	0	10.95	540.56	80.5	1500	9929	220	208	0	0	0	0	0	TN	TN
6500	0	0.6	8207.45	79.54	1500	9929	220	208	0	0	0	0	0	TN	TN
6502	0	0.85	1339.1	78.51	1500	9929	220	208	0	0	0	0	0	TN	TN
6504	0	1.27	924.78	61.58	1500	9929	220	208	0	0	0	0	0	TN	TN
6509	0	9.47	660.68	106.9	1500	9929	220	208	0	0	0	0	0	TN	TN
6514	0	0.38	509.65	138.29	1500	9929	220	208	0	0	0	0	0	TN	TN
6522	0	21.38	1334.45	56.91	1500	9929	220	208	0	0	0	0	0	TN	TN
6542	0	6.06	456.64	97.56	1500	9929	220	208	0	0	0	0	0	TN	TN
6565	0	11.74	543.85	46.06	1500	9929	220	208	0	0	0	0	0	TN	TN
6573	0	7.77	447.58	98.62	1500	9929	220	208	0	0	0	0	0	TN	TN
6523	0	292.98	304540.9	6835.76	1500	9929	220	208	0	1	1	1	1	FP	FP
6524	0	19.51	646.99	91.27	1500	9929	220	208	0	0	0	0	0	TN	TN
6527	0	32.39	1542.39	92.6	1500	9929	220	208	0	0	0	0	0	TN	TN
6529	0	13.33	2896.03	57.63	1500	9929	220	208	0	0	0	0	0	TN	TN
6543	0	8.73	568.31	39.1	1500	9929	220	208	0	0	0	0	0	TN	TN
6554	0	27.46	1310.09	59.7	1500	9929	220	208	0	0	0	0	0	TN	TN
6585	0	112.35	1095.98	90.79	1500	9929	220	208	0	0	0	0	0	TN	TN
6586	0	28.9	605.76	44.78	1500	9929	220	208	0	0	0	0	0	TN	TN
6589	0	15.09	1444.35	41.85	1500	9929	220	208	0	0	0	0	0	TN	TN

Table 3A.3 Cont'd

6590	0	6.85	777.23	92.17	1500	9929	220	208	0	0	0	0	0	TN	TN
6592	0	5.69	281.19	127.05	1500	9929	220	208	0	0	0	0	0	TN	TN
6604	0	51.97	2688.23	116.38	1500	9929	220	208	0	0	0	0	0	TN	TN
6605	0	118.69	3038.16	154.31	1500	9929	220	208	0	0	0	0	0	TN	TN
6614	0	5.56	608.25	43.99	1500	9929	220	208	0	0	0	0	0	TN	TN
6617	0	5.78	283.51	95.77	1500	9929	220	208	0	0	0	0	0	TN	TN
6532	0	6.29	360.23	29.29	1500	9929	220	208	0	0	0	0	0	TN	TN
6537	0	6.74	149.35	155.13	1500	9929	220	208	0	0	0	0	0	TN	TN
6539	0	6.91	54.52	57.5	1500	9929	220	208	0	0	0	0	0	TN	TN
6555	0	15.7	860.81	42.71	1500	9929	220	208	0	0	0	0	0	TN	TN
6557	0	8.42	412.37	119.45	1500	9929	220	208	0	0	0	0	0	TN	TN
6567	0	6.7	325.38	75.38	1500	9929	220	208	0	0	0	0	0	TN	TN
6574	0	13.8	183.54	90.08	1500	9929	220	208	0	0	0	0	0	TN	TN
6578	0	6.88	269.65	79.18	1500	9929	220	208	0	0	0	0	0	TN	TN
6595	0	5.61	757.39	134.65	1500	9929	220	208	0	0	0	0	0	TN	TN

\*FN, false negative; TP, true positive; FP, false positive; TN, true negative.

0=benign (Control), 1=PDAC (Case)

Table 3A.4 Application of the sensitivity panel and CA19-9 to blinded samples.

		Marker values			Preset panel thresholds		Preset CA19-9 alone threshold	Blinded panel markers, thresholded			Blinded panel call	Blinded CA19-9 call	Blinded results	
ID	Status (0=benign 1=PDAC)	CA19-9	CA19-9: sTRA	MUC16: sTRA	CA19-9: sTRA	MUC16: sTRA	CA19-9	CA19-9: sTRA	MUC16: TRA	1=case 0=control	1=case 0=control	Panel	CA19-9	
6437	1	0.51	469.77	61.63	470	100	6	0	0	0	0	FN	FN	
6438	1	14.04	1743.82	31.71	470	100	6	1	0	1	1	TP	TP	
6440	1	24.74	1777.63	40.37	470	100	6	1	0	1	1	TP	TP	
6441	1	25.29	1530.89	22.06	470	100	6	1	0	1	1	TP	TP	
6444	1	207.72	161706.3	29.48	470	100	6	1	0	1	1	TP	TP	
6446	1	151.35	12864.24	81.42	470	100	6	1	0	1	1	TP	TP	
6447	1	48.49	3779.54	36.87	470	100	6	1	0	1	1	TP	TP	
6451	1	1001.8	199156.9	23.5	470	100	6	1	0	1	1	TP	TP	
6453	1	629.34	62686.21	26.15	470	100	6	1	0	1	1	TP	TP	
6456	1	57.23	20587.6	47.37	470	100	6	1	0	1	1	TP	TP	
6458	1	33.63	111672.9	6162.7	470	100	6	1	1	1	1	TP	TP	
6459	1	44.98	4914.95	80.35	470	100	6	1	0	1	1	TP	TP	
6460	1	257.66	23806.19	31.92	470	100	6	1	0	1	1	TP	TP	
6464	1	309.58	24258.2	34.69	470	100	6	1	0	1	1	TP	TP	
6466	1	254.47	32267.11	12.75	470	100	6	1	0	1	1	TP	TP	
6469	1	106.09	10481.75	33.05	470	100	6	1	0	1	1	TP	TP	
6472	1	704.18	40180.51	34.82	470	100	6	1	0	1	1	TP	TP	
6473	1	635.92	29143.2	103.51	470	100	6	1	1	1	1	TP	TP	
6474	1	5.01	471.37	41.29	470	100	6	1	0	1	0	TP	FN	
6475	1	26.34	7865.06	48.42	470	100	6	1	0	1	1	TP	TP	
6476	1	100.08	25170.23	30.2	470	100	6	1	0	1	1	TP	TP	
6478	1	24.14	1111.44	54.52	470	100	6	1	0	1	1	TP	TP	
6481	1	58.57	19827.56	19.92	470	100	6	1	0	1	1	TP	TP	
6482	1	306.86	3827.78	27.51	470	100	6	1	0	1	1	TP	TP	
6484	1	84.78	23539.76	68.58	470	100	6	1	0	1	1	TP	TP	
6485	1	0.44	1866.3	34.97	470	100	6	1	0	1	0	TP	FN	
6489	1	7561.44	652598.6	69.34	470	100	6	1	0	1	1	TP	TP	
6490	1	33.47	9651.47	41.21	470	100	6	1	0	1	1	TP	TP	
6491	1	508.03	45991.96	58.15	470	100	6	1	0	1	1	TP	TP	

Table 3A.4 Cont'd

6492	1	22.43	8901.39	37.7	470	100	6	1	0	1	1	TP	TP
6494	1	167.3	5771.59	61.96	470	100	6	1	0	1	1	TP	TP
6496	1	2206.45	346997.5	7310.52	470	100	6	1	1	1	1	TP	TP
6497	1	14543.3	504125	112.49	470	100	6	1	1	1	1	TP	TP
6503	1	148.38	4603.99	42.26	470	100	6	1	0	1	1	TP	TP
6505	1	67.52	5936.85	92.93	470	100	6	1	0	1	1	TP	TP
6506	1	1.81	923.41	41.19	470	100	6	1	0	1	0	TP	FN
6507	1	38.43	1071.31	52.28	470	100	6	1	0	1	1	TP	TP
6508	1	1.27	169.98	38.13	470	100	6	0	0	0	0	FN	FN
6510	1	0.44	5932.9	167.91	470	100	6	1	1	1	0	TP	FN
6515	1	2724.86	89836.24	242.57	470	100	6	1	1	1	1	TP	TP
6517	1	146.68	964.77	59.47	470	100	6	1	0	1	1	TP	TP
6526	1	10.36	2497.84	88.95	470	100	6	1	0	1	1	TP	TP
6531	1	230.71	66730.45	91.95	470	100	6	1	0	1	1	TP	TP
6540	1	173.95	35766.54	81.23	470	100	6	1	0	1	1	TP	TP
6541	1	141.15	13217.11	37.96	470	100	6	1	0	1	1	TP	TP
6544	1	132.12	134067.7	189.08	470	100	6	1	1	1	1	TP	TP
6545	1	582.93	137330.3	105.04	470	100	6	1	1	1	1	TP	TP
6547	1	5.97	529.8	37	470	100	6	1	0	1	0	TP	FN
6550	1	260.31	4636.38	56.81	470	100	6	1	0	1	1	TP	TP
6551	1	8.18	1067.06	37.58	470	100	6	1	0	1	1	TP	TP
6558	1	104.54	35703.93	59.17	470	100	6	1	0	1	1	TP	TP
6561	1	173.93	64733.53	241.55	470	100	6	1	1	1	1	TP	TP
6562	1	112.04	16640.75	42.6	470	100	6	1	0	1	1	TP	TP
6564	1	1495.94	94374.81	74.16	470	100	6	1	0	1	1	TP	TP
6571	1	212.5	80906.49	56.36	470	100	6	1	0	1	1	TP	TP
6572	1	50.6	10747.94	50.54	470	100	6	1	0	1	1	TP	TP
6576	1	576.79	18532.79	151.89	470	100	6	1	1	1	1	TP	TP
6577	1	16.69	4227.35	27.47	470	100	6	1	0	1	1	TP	TP
6580	1	65.38	9897.59	250.84	470	100	6	1	1	1	1	TP	TP
6583	1	5.62	251.08	83.93	470	100	6	0	0	0	0	FN	FN
6587	1	158.49	35792.78	4604.86	470	100	6	1	1	1	1	TP	TP
6597	1	124.96	2781.09	63.74	470	100	6	1	0	1	1	TP	TP
6598	1	62.84	20628.93	153.06	470	100	6	1	1	1	1	TP	TP
6600	1	99.39	7586.51	58.27	470	100	6	1	0	1	1	TP	TP

Table 3A.4 Cont'd

6600	1	99.39	7586.51	58.27	470	100	6	1	0	1	1	TP	TP
6601	1	1106.73	551586.2	8786.2	470	100	6	1	1	1	1	TP	TP
6602	1	333.57	391359.7	1509.34	470	100	6	1	1	1	1	TP	TP
6603	1	129.18	32717.41	91.17	470	100	6	1	0	1	1	TP	TP
6607	1	5.04	430.56	71.95	470	100	6	0	0	0	0	FN	FN
6609	1	7.08	109.36	1170.68	470	100	6	0	1	1	1	TP	TP
6612	1	90.9	588.99	71.48	470	100	6	1	0	1	1	TP	TP
6616	1	7.24	398.73	117.14	470	100	6	0	1	1	1	TP	TP
6518	0	8.51	339.49	51.93	470	100	6	0	0	0	1	TN	FP
6520	0	23.8	887.54	33.98	470	100	6	1	0	1	1	FP	FP
6521	0	5.51	482.85	51.48	470	100	6	1	0	1	0	FP	TN
6525	0	8.37	240.04	26.62	470	100	6	0	0	0	1	TN	FP
6533	0	15.32	1637.83	66.49	470	100	6	1	0	1	1	FP	FP
6534	0	5.58	795.71	20.77	470	100	6	1	0	1	0	FP	TN
6535	0	5.74	354.79	85.09	470	100	6	0	0	0	0	TN	TN
6536	0	12.89	651.06	135.82	470	100	6	1	1	1	1	FP	FP
6556	0	6.95	224.03	59.56	470	100	6	0	0	0	1	TN	FP
6559	0	13.35	218.81	76.57	470	100	6	0	0	0	1	TN	FP
6569	0	11.35	357	67.74	470	100	6	0	0	0	1	TN	FP
6570	0	6.83	1028.47	100.27	470	100	6	1	1	1	1	FP	FP
6575	0	6.6	368.93	53.24	470	100	6	0	0	0	1	TN	FP
6579	0	9.7	241.32	76.04	470	100	6	0	0	0	1	TN	FP
6582	0	5.53	421.24	72.73	470	100	6	0	0	0	0	TN	TN
6584	0	8.27	252.54	46.77	470	100	6	0	0	0	1	TN	FP
6588	0	6.21	204.23	36.24	470	100	6	0	0	0	1	TN	FP
6594	0	8.31	6965.61	103.6	470	100	6	1	1	1	1	FP	FP
6608	0	6.16	572.5	40.27	470	100	6	1	0	1	1	FP	FP
6615	0	5.26	378.69	11.66	470	100	6	0	0	0	0	TN	TN
6528	0	27.02	2475.54	135.1	470	100	6	1	1	1	1	FP	FP
6548	0	5.26	120.42	40.66	470	100	6	0	0	0	0	TN	TN
6553	0	20.64	2048.46	71.77	470	100	6	1	0	1	1	FP	FP
6563	0	6.46	201.82	34.68	470	100	6	0	0	0	1	TN	FP
6566	0	15.9	444.4	64.23	470	100	6	0	0	0	1	TN	FP
6591	0	16.26	3600.64	98.62	470	100	6	1	0	1	1	FP	FP
6599	0	3181.49	12401.13	62.22	470	100	6	1	0	1	1	FP	FP

Table 3A.4 Cont'd

6613	0	7.56	754.58	49.31	470	100	6	1	0	1	1	FP	FP
6439	0	0.54	388.13	106.29	470	100	6	0	1	1	0	FP	TN
6443	0	5.21	602.88	40.24	470	100	6	1	0	1	0	FP	TN
6445	0	2.02	617.89	26.67	470	100	6	1	0	1	0	FP	TN
6449	0	1.79	1119.3	31.3	470	100	6	1	0	1	0	FP	TN
6450	0	2.8	442.07	36.49	470	100	6	0	0	0	0	TN	TN
6455	0	4.33	696.99	27.84	470	100	6	1	0	1	0	FP	TN
6462	0	8.43	1722.52	119.78	470	100	6	1	1	1	1	FP	FP
6463	0	1.91	14674.27	57.31	470	100	6	1	0	1	0	FP	TN
6465	0	0.44	300.26	25.04	470	100	6	0	0	0	0	TN	TN
6471	0	0.38	645.39	54.64	470	100	6	1	0	1	0	FP	TN
6483	0	1.25	177.05	29.32	470	100	6	0	0	0	0	TN	TN
6486	0	0.34	388.55	46.46	470	100	6	0	0	0	0	TN	TN
6487	0	2.84	335.6	16.46	470	100	6	0	0	0	0	TN	TN
6493	0	12.33	661.67	94.49	470	100	6	1	0	1	1	FP	FP
6498	0	10.95	540.56	44.57	470	100	6	1	0	1	1	FP	FP
6500	0	0.6	8207.45	150.21	470	100	6	1	1	1	0	FP	TN
6502	0	0.85	1339.1	71.07	470	100	6	1	0	1	0	FP	TN
6504	0	1.27	924.78	40.81	470	100	6	1	0	1	0	FP	TN
6509	0	9.47	660.68	60.69	470	100	6	1	0	1	1	FP	FP
6514	0	0.38	509.65	18.16	470	100	6	1	0	1	0	FP	TN
6522	0	21.38	1334.45	73.8	470	100	6	1	0	1	1	FP	FP
6542	0	6.06	456.64	55	470	100	6	0	0	0	1	TN	FP
6565	0	11.74	543.85	25.38	470	100	6	1	0	1	1	FP	FP
6573	0	7.77	447.58	28.63	470	100	6	0	0	0	1	TN	FP
6523	0	292.98	304540.9	3416.25	470	100	6	1	1	1	1	FP	FP
6524	0	19.51	646.99	96.56	470	100	6	1	0	1	1	FP	FP
6527	0	32.39	1542.39	38.81	470	100	6	1	0	1	1	FP	FP
6529	0	13.33	2896.03	57.85	470	100	6	1	0	1	1	FP	FP
6543	0	8.73	568.31	67.53	470	100	6	1	0	1	1	FP	FP
6554	0	27.46	1310.09	46.81	470	100	6	1	0	1	1	FP	FP
6585	0	112.35	1095.98	69.86	470	100	6	1	0	1	1	FP	FP
6586	0	28.9	605.76	64.61	470	100	6	1	0	1	1	FP	FP
6589	0	15.09	1444.35	92.42	470	100	6	1	0	1	1	FP	FP
6590	0	6.85	777.23	30.47	470	100	6	1	0	1	1	FP	FP

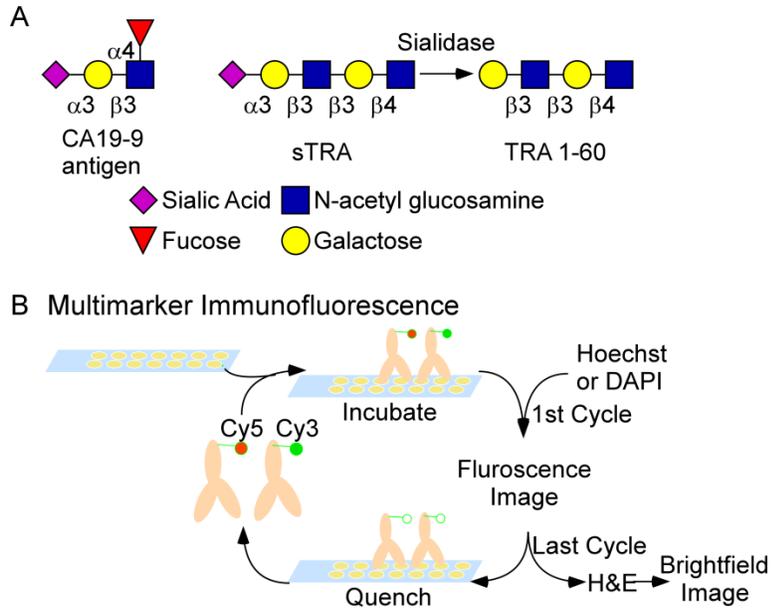
Table 3A.4 Cont'd

6592	0	5.69	281.19	101.74	470	100	6	0	1	1	0	FP	TN
6604	0	51.97	2688.23	152.47	470	100	6	1	1	1	1	FP	FP
6605	0	118.69	3038.16	86.28	470	100	6	1	0	1	1	FP	FP
6614	0	5.56	608.25	681.14	470	100	6	1	1	1	0	FP	TN
6617	0	5.78	283.51	53.14	470	100	6	0	0	0	0	TN	TN
6532	0	6.29	360.23	73.42	470	100	6	0	0	0	1	TN	FP
6537	0	6.74	149.35	91.02	470	100	6	0	0	0	1	TN	FP
6539	0	6.91	54.52	64.88	470	100	6	0	0	0	1	TN	FP
6555	0	15.7	860.81	41.96	470	100	6	1	0	1	1	FP	FP
6557	0	8.42	412.37	65.22	470	100	6	0	0	0	1	TN	FP
6567	0	6.7	325.38	29.86	470	100	6	0	0	0	1	TN	FP
6574	0	13.8	183.54	109.18	470	100	6	0	1	1	1	FP	FP
6578	0	6.88	269.65	30.96	470	100	6	0	0	0	1	TN	FP
6595	0	5.61	757.39	124.18	470	100	6	1	1	1	0	FP	TN

\*FN, false negative; TP, true positive; FP, false positive; TN, true negative.

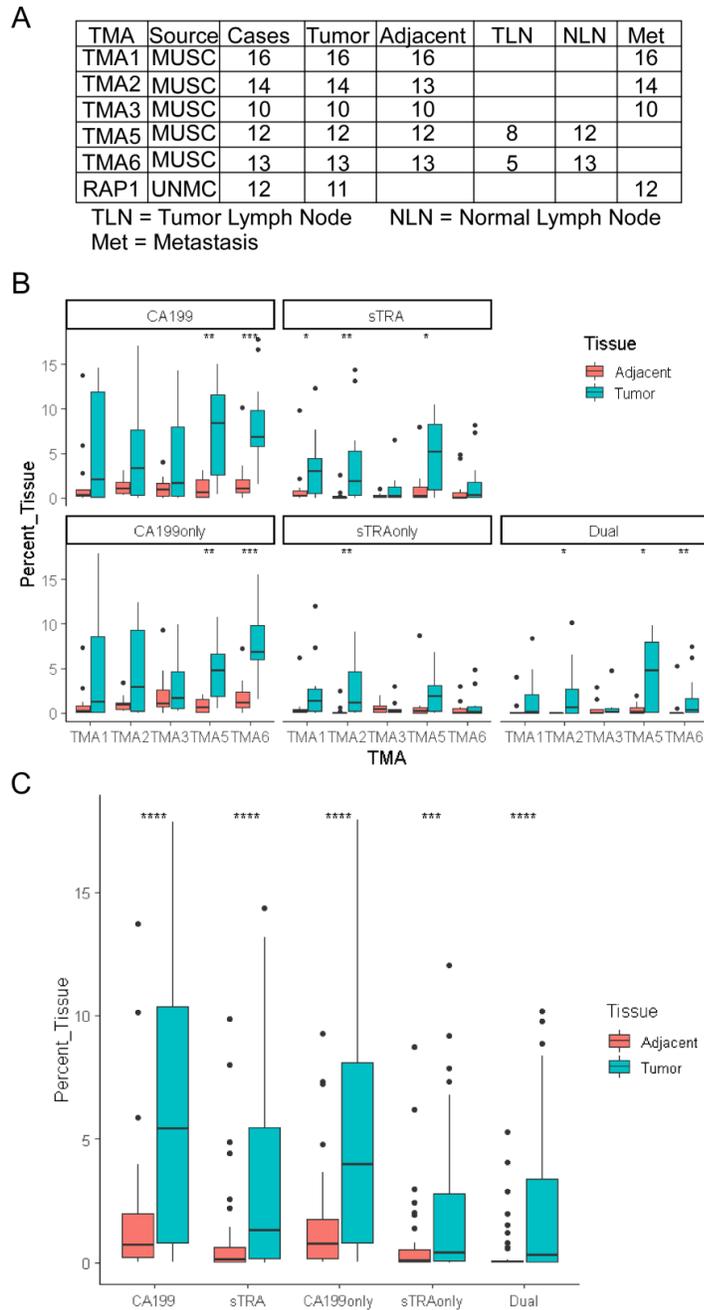
## Chapter 4 Figures

Figure 4.1 The CA19-9 and sTRA glycans and their detection on tissue microarrays



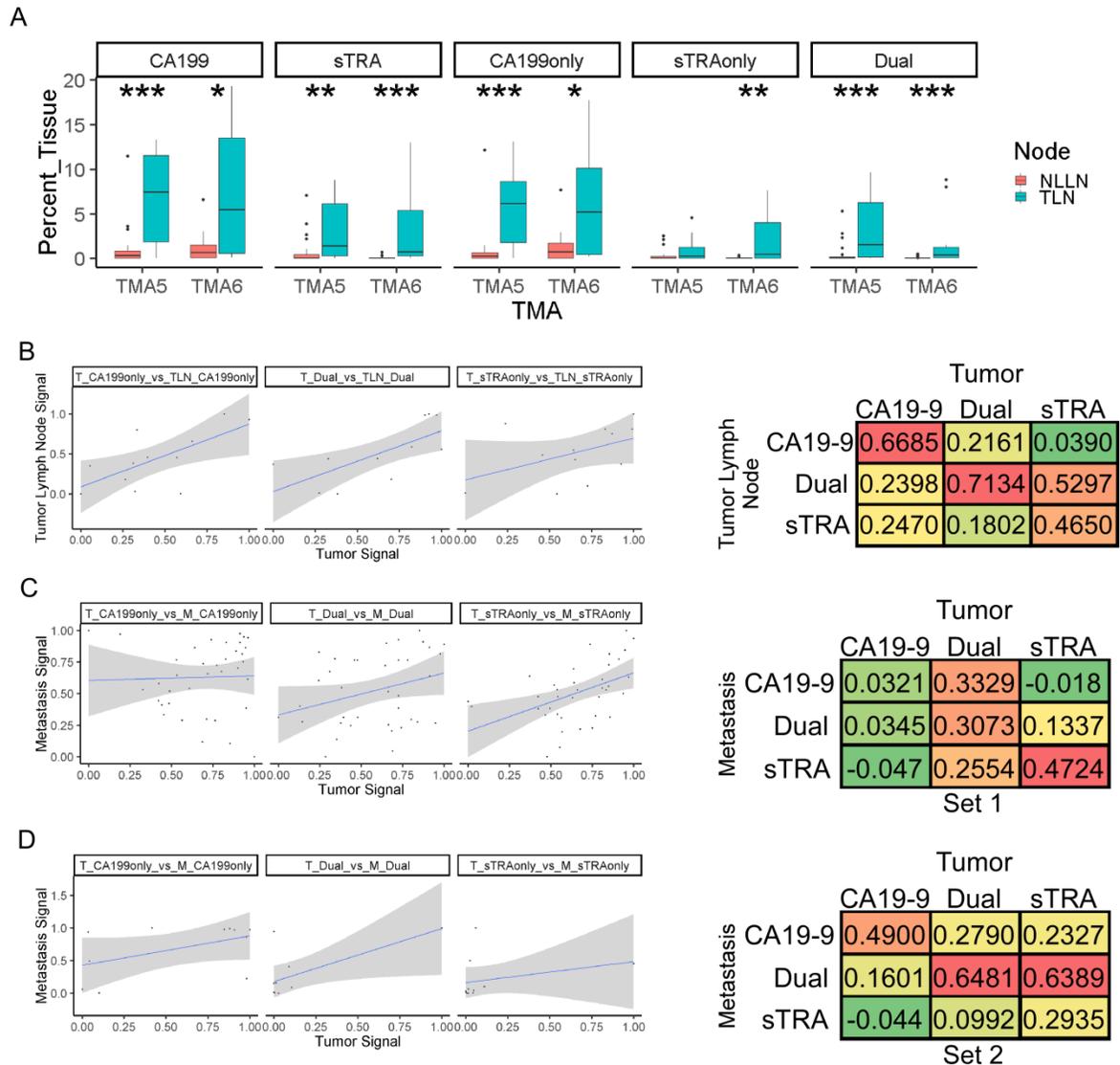
A. The CA19-9 and sTRA glycans in representative structural form. B. Multimarker immunofluorescence workflow shows the detection with cy5 and cy3 conjugated antibodies in successive rounds followed by H&E at the end of immunofluorescence used for multidetection of antigens on tissues.

Figure 4.2 Validation of the CA19-9 and sTRA determined glycotypes.



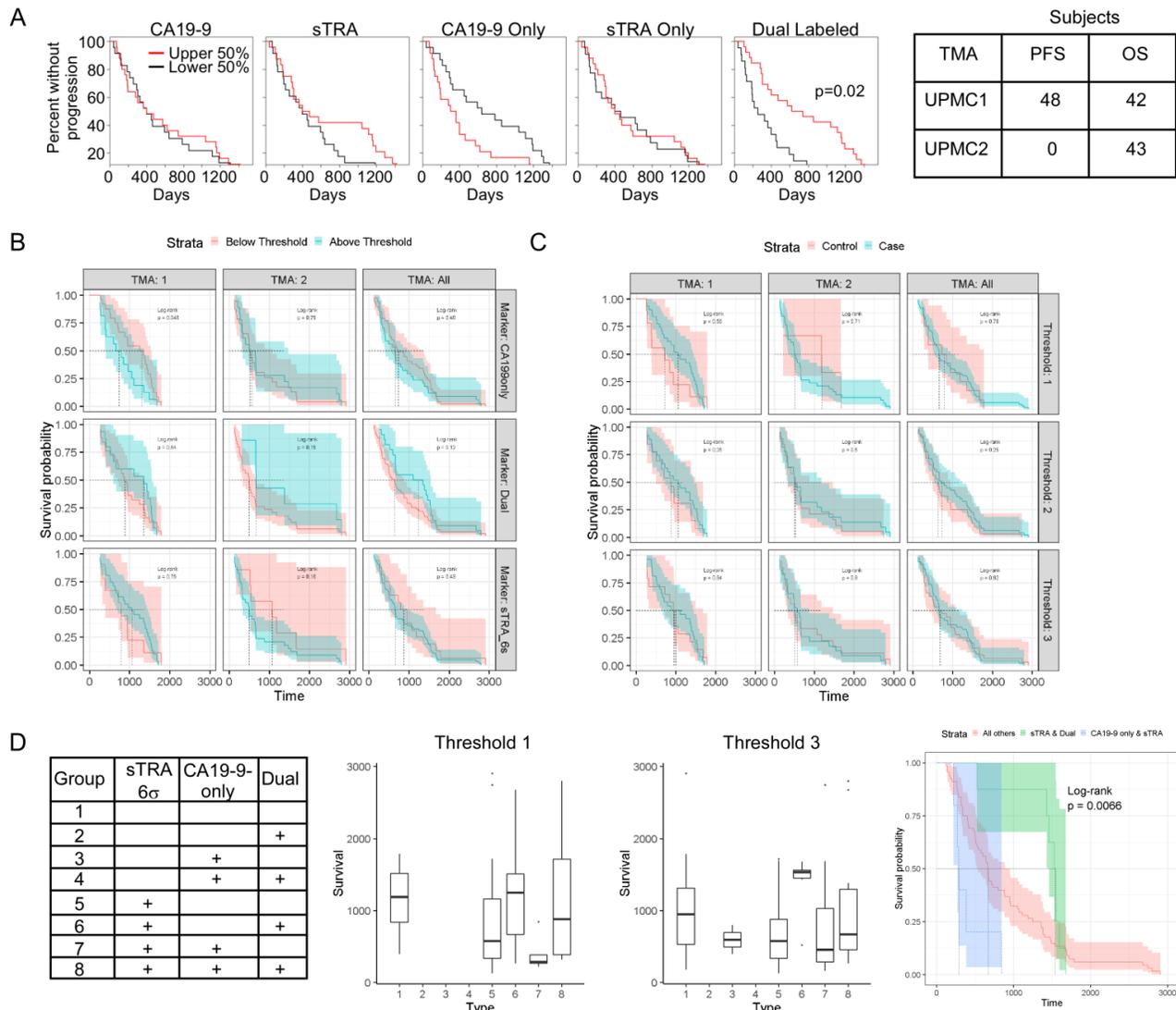
A. For tumor and metastasis analysis, we used 6 TMAs consisting of 77 cases with 64 subjects with matched tumor and adjacent tissues (2 cores per patient), 13 subjects with matched tumor, tumor lymph nodes (TLN) and normal lymph nodes (NLN), 25 total subjects with normal lymph nodes, and 40 subjects with matched tumors and metastases from one side and 11 matched tumors and metastases with larger 5mm cores from an independent site. B. Aggregate analysis of sTRA-only, CA19-9 only, dual expression, and all CA19-9 and all sTRA showing significant associations between tumor and adjacent tissues (\*\*\*\* $p < 0.0001$ ).

Figure 4.3 Tumor lymph nodes and metastases show correlation with their origin tumors.



A. Tumor lymph node vs normal lymph node, each of the glycotypes (CA19-9 only, sTRA only, dual (CA19-9/sTRA) and the overall population expression for total CA19-9 and total sTRA) shows significant elevation over normal lymph node controls ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ). B. Cross correlations of normalized data (normalization testing in Supplemental Figure 4A.1) for tumor vs tumor lymph nodes for CA19-9 only, sTRA only and dual CA19-9/sTRA expression showing strongest correlations within each marker from tumor (T) to tumor lymph node (TLN). The gray bands around the regression represent the 95% CI (B,C, and D). C. Distant metastasis with correlation to baseline tumor biomarkers. TMAs1-3 aggregate cross correlation analysis for CA19-9 only, sTRA only, and dual expression. sTRA expression showed the strongest correlation between tumor and metastasis. Additional correlation combinations are in Figure 4A.3. D. A small tumor-matched metastasis validation set from UNMC performs inconsistently with the previous aggregated TMA set for distant metastases, but aligned fairly well with the correlation patterns for lymph nodes. B-D. Pearson's correlation grids for the cross correlations shown in the scatterplots adjacent show correlations consistent with the sample distributions adjacent to the grid.

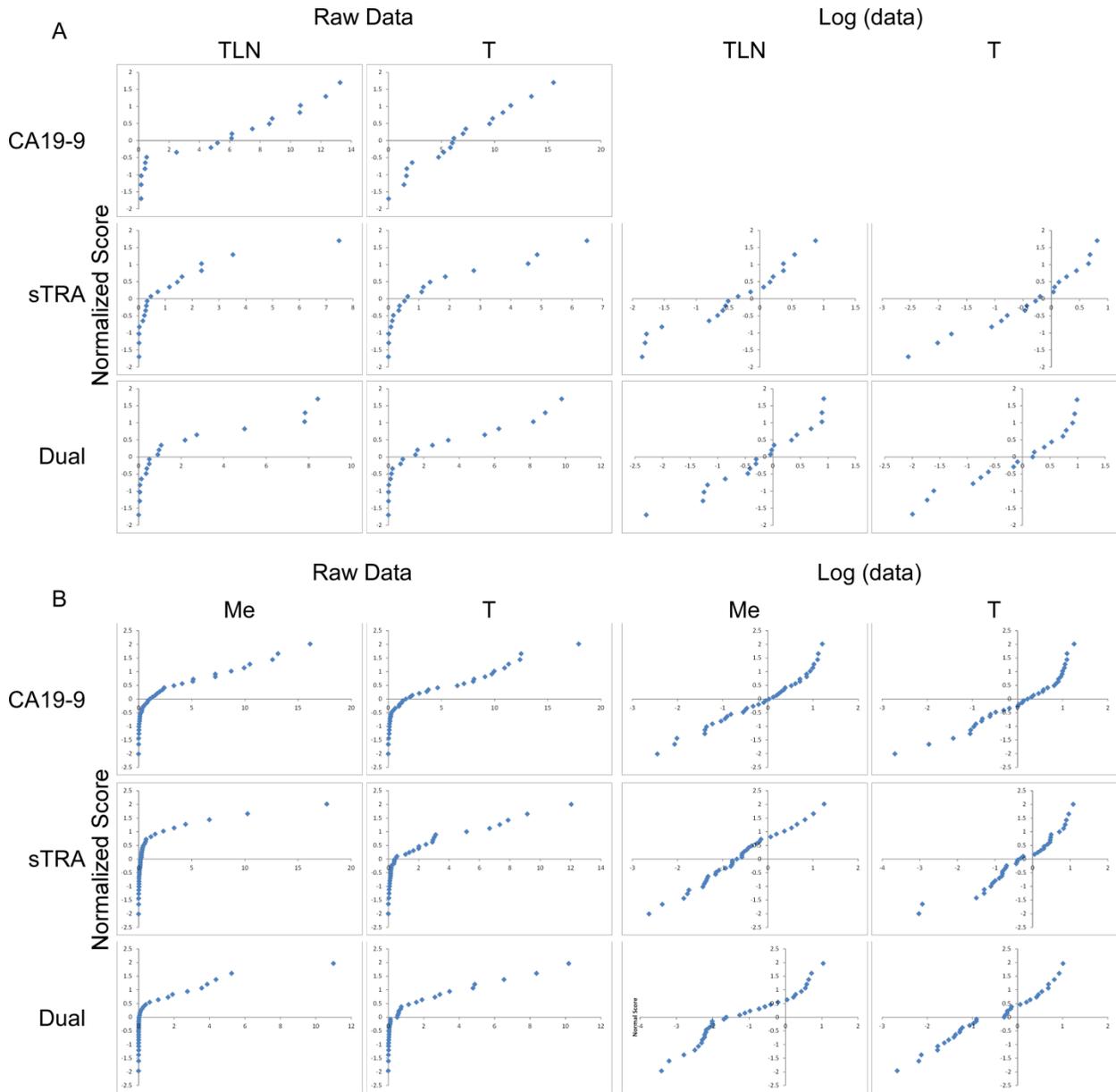
Figure 4.4 Survival analysis shows weak anticorrelation for the upper and lower ends of dual expression.



A. Progression-free survival analysis shows improved time to progression for dual glycan-expressing tumors (significance testing by logrank). A table shows 43 and 42 patients tissue microarrays 1 and 2, respectively, with overall survival for analysis. B. We analyzed a second tissue microarray independent from our preliminary data and examined overall survival on both tissue microarrays. Two group stratification (high/low) fell short of validation. Data is shown for markers above and below thresholds for optimized panel thresholds. C. Three threshold data is shown by TMA or combined group. Threshold 1 was set by progression-free survival optimization using MSS thresholding on TMA 1. Overall survival threshold was determined for TMA 2, and an average of thresholds 1 and 2 was used for threshold three. D. Using the thresholds of the three markers determined to contribute most to survival differences by MSS optimization, there are 8 possible groups (left panel). The distribution of these groups is shown for thresholds 1 and 3 and the survival of Groups 6 (sTRA+/Dual+) and 7 (sTRA+/CA19-9 only+) was plotted against all non-Groups 6 and 7 for survival differences. Group 6 shows longer survival and Group 7 shows short survival. Log rank is  $p < 0.01$  for differences between the groups, bands represent 95% confidence interval.

## Chapter 4 Supplemental Figures

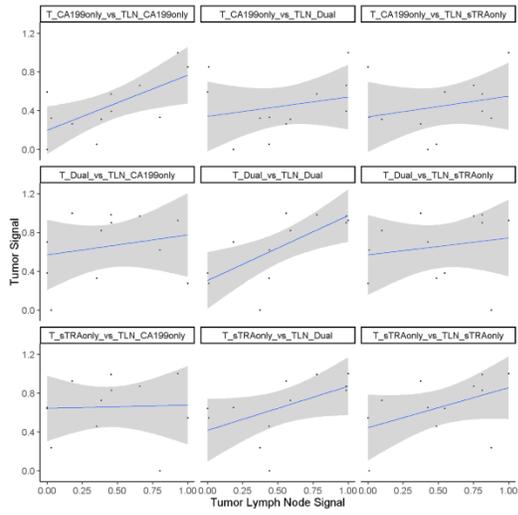
Figure 4A.1 Normal Distribution testing for metastatic data with normalization



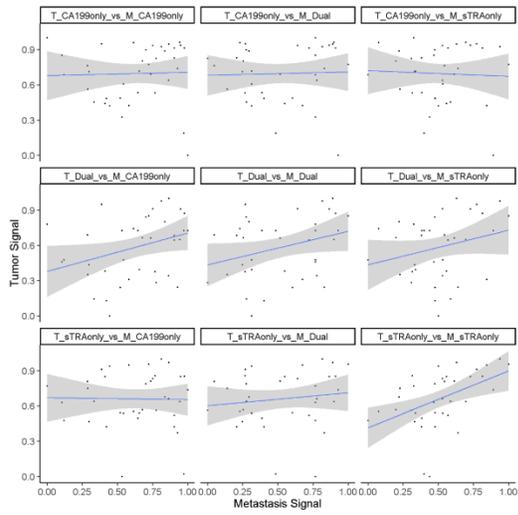
A. Tumor lymph node biomarker distribution shows log normal distribution for all markers except CA19-9. Log-normal corrections are shown at right. B. Metastasis biomarker distribution shows log normal distribution in all biomarkers with log normal distribution normalization at the right.

Figure 4A.2 Regional and distant metastasis glycotype correlations

A



B



C

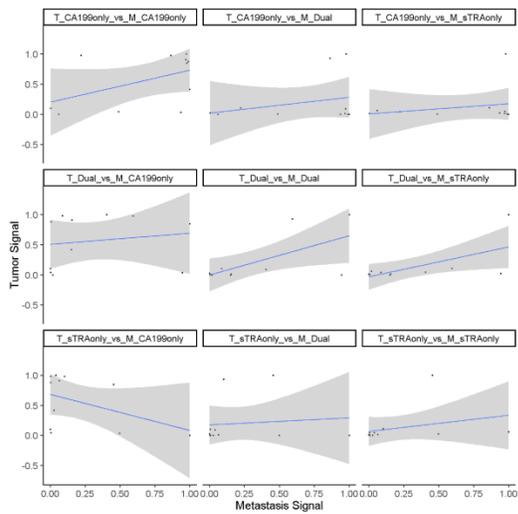
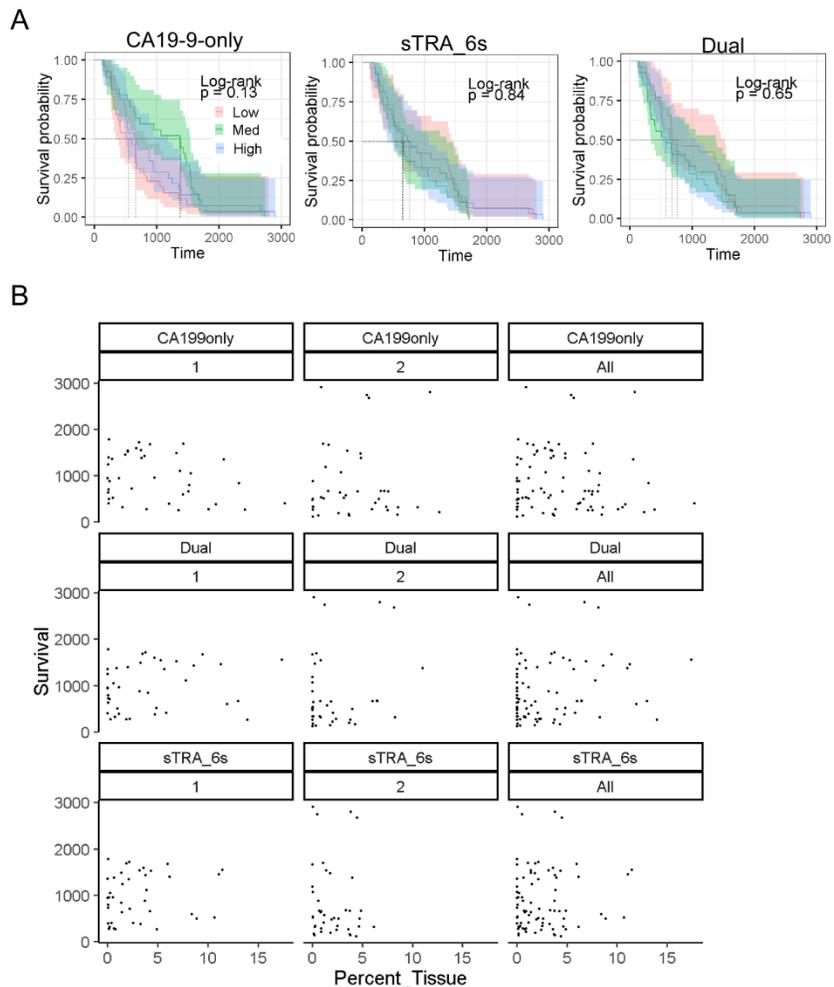


Figure 4A.2 Cont'd

A. Tumor vs Tumor lymph node glycotype correlations show strong the strongest correlations between the same glycotype (i.e. CA19-9 tumor vs CA19-9 lymph node), with notable lack of correlation between CA19-9 and sTRA from tumor to tumor lymph node. Tumor sTRA also shows a moderate correlation to dual expression in tumor lymph nodes. B. Set 1 Tumor vs Metastasis glycotype correlation shows strong correlation between sTRA and dual in tumor with the same glycotype in distant metastases. Notable lack of correlation between CA19-9 and sTRA as in lymph node metastases. C. Set 2 Tumor vs Metastasis demonstrates a similar pattern to lymph nodes. Linear regression calculated by least squares with 95% confidence interval.

Figure 4A.3 Survival distribution and association with glycan expression



A. Tertile distributions of the three markers used in panel analysis do not show significant differences between expression and survival. B. The absolute value of glycan expression is not a significant determinant of survival. Thresholding and glycan combination have more significant impact on survival.

Table 4A.1 Demographic data for patients on clinical tissue microarrays

TMA	Race	Sex	Age	Vital Status	Recurrence	Type of Recurrence	Histological Type	Histological Grade	Staging
TMA1	WHITE	M	75		Unknown		Adeno-carcinoma	moderately differentiated	T3N0MX
TMA1	WHITE	M	76		Unknown		Cholangio-carcinoma	moderately differentiated	T3N1MX
TMA1	WHITE	F	61		YES	distant recurrence - GAST	Adeno-carcinoma	well differentiated	T3N0MX
TMA1	WHITE	M	73		YES	Distant recurrence - CNS	Mucinous Adeno-carcinoma		
TMA1	WHITE	M	80		Unknown		Adeno-carcinoma	poorly differentiated	T3N1MX
TMA1	WHITE	F	69		Unknown		Adeno-carcinoma	moderately differentiated	T3N0MX
TMA1	BLACK	M	73		Unknown		Adeno-carcinoma	well differentiated	T3N1MX
TMA1	BLACK	F	61			metastacized to liver and omentum	Adeno-carcinoma	poorly differentiated	T4N1Mx
TMA1	WHITE	M	65		Unknown		Adeno-carcinoma	well differentiated	T3N1MX
TMA1	BLACK	F	67		Unknown		Adeno-carcinoma	poorly differentiated	T3N1MX
TMA1	WHITE	M	69		Unknown		Adeno-carcinoma		T3N1MX
TMA1	WHITE	M	71			Never disease free	Adeno-carcinoma	poorly differentiated	T3N1MX
TMA1	WHITE	M	83			Disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA1	WHITE	F	63			Possible recurrence	Adeno-carcinoma	poorly differentiated	T3N1MX
TMA1	WHITE	M	78			Disease free	Adeno-carcinoma	moderately differentiated	T3N1
TMA1	BLACK	M	66			Never disease free	Adeno-carcinoma	poorly differentiated	T3N1MX
TMA2	WHITE	F	58			Disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA2	WHITE	M	69		Unknown		Adeno-carcinoma	poorly differentiated	T3NXMX
TMA2	BLACK	F	51		Unknown		Adeno-carcinoma	moderately differentiated	T3N1MX
TMA2	WHITE	M	42		Unknown		Adeno-carcinoma	poorly differentiated	T3N1MX
TMA2	WHITE	F	66		Unknown		Adeno-carcinoma	moderately differentiated	T2N1MX

Table 4A.1 Cont'd

TMA2	WHITE	F	87			Disease free	Adeno-carcinoma	moderately differentiated	T1NOMX
TMA2	WHITE	F	54		YES	Local Recurrence	Adeno-carcinoma		T2NOMX
TMA2	WHITE	M	56	Dead		Never disease free	Adeno-carcinoma		T3N1MX
TMA2	BLACK	F	62			Never disease free	Adeno-carcinoma	poorly differentiated	T3NOMX
TMA2	WHITE	F	76				Ductal adeno-carcinoma	moderately differentiated	T3NOMX
TMA2	BLACK	F	68		YES	Local recurrence	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA2	WHITE	F	37		YES	Distant recurrence	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA2	WHITE	M	78			Never disease free	Adeno-carcinoma	moderately differentiated	T2N1bMX
TMA2	WHITE	M	70			Never disease free	Adeno-carcinoma		T2NOMX
TMA3	BLACK	M	48			Never disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	BLACK	M	62		YES	Local recurrence	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	F	69			Never disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	F	82			Never disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	F	77			Disease free	Adeno-carcinoma	poorly differentiated	T3NOMX
TMA3	WHITE	M	66			Disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	F	51		YES	Distant recurrence - lung	Ductal adeno-carcinoma	moderately differentiated	T2NOMX
TMA3	WHITE	M	75			Disease free	Cholangio-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	M	62			Never disease free	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA3	WHITE	F	82			Disease free	Adeno-carcinoma	moderately differentiated	T3NOMX
TMA5	WHITE	M	71		YES	Distant Reurrence - Lung	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA5	WHITE	F	50	DEAD	YES	Distant recurrence - Hept	Ductal adeno-carcinoma	moderately differentiated	T3N1MX

Table 4A.1 Cont'd

TMA5	WHITE	F	53	DEAD	YES	Distant recurrence - Lung	Ductal adenocarcinoma	moderately differentiated	T2N0MX
TMA5	WHITE	F	62		Unknown		Ductal adenocarcinoma	moderately differentiated	T3N1bM
TMA5	WHITE	M	62		YES	Distant recurrence - Hept	Adenocarcinoma	moderately differentiated	T3N1bMX
TMA5	WHITE	F	71			Disease Free	Ductal adenocarcinoma	moderately differentiated	T3N0MX
TMA5	WHITE	F	60			Disease Free	Ductal adenocarcinoma	moderately differentiated	T3N1MX
TMA5	BLACK	M	64			Disease Free	Ductal adenocarcinoma	moderately differentiated	T1N1bMX
TMA5	WHITE	F	78			Disease Free	Adenocarcinoma	moderately differentiated	T1N0MX
TMA5	WHITE	M	87			Disease Free	Adenocarcinoma	poorly differentiated	T3N1bMX
TMA5	BLACK	M	51		YES	Distant recurrence - Hept	Ductal adenocarcinoma	moderately differentiated	T3N1MX
TMA5	ASIAN	M	39	DEAD		Never disease free	Ductal adenocarcinoma	moderately differentiated	T3N1MX
TMA6	WHITE	M	73		YES	Distant recurrence - Lung	Adenocarcinoma	moderately differentiated	T3N0MX
TMA6	WHITE	F	71		YES	Local recurrence	Adenocarcinoma	moderately differentiated	T3N1MX
TMA6	WHITE	F	64			Disease Free	Adenocarcinoma		T2NXMX
TMA6	WHITE	M	71		YES	Distant recurrence - Hept	Adenocarcinoma	moderately differentiated	T3N0MX
TMA6	WHITE	M	69		YES	Distant recurrence - Adrenal Gland	Ductal adenocarcinoma	moderately differentiated	T3N1MX
TMA6	WHITE	M	65			Disease Free	Ductal adenocarcinoma	moderately differentiated	T3N1MX

Table 4A.1 Cont'd

TMA6	BLACK	M	74		YES	Distant recurrence - Hept	Adeno-carcinoma	moderately differentiated	T3N1MX
TMA6	WHITE	F	86			Disease Free	Ductal adeno-carcinoma	moderately differentiated	T1N0NX
TMA6	BLACK	M	73		YES	Local/Distant recurrence - Panc&Gast	Ductal adeno-carcinoma	moderately differentiated	T3N1
TMA6	WHITE	M	47			Disease Free	Ductal adeno-carcinoma	moderately differentiated	T3N0MX
TMA6	WHITE	F	65			Disease Free	Ductal adeno-carcinoma	moderately differentiated	T3N1
TMA6	WHITE	F	66		YES	Distant recurrence – Ovar	Mixed Colloid/Ductal Adeno-carcinoma	well differentiated	T3N0MX
TMA6	WHITE	M	77		YES	Local Recurrence	Adeno-carcinoma	poorly differentiated	T3N1MX

## Supplementary Methods

Automated identification and quantification of signals in multichannel immunofluorescence images: the SignalFinder platform

**Daniel Barnett,\* Johnathan Hall,\* and Brian Haab**

Van Andel Research Institute, Grand Rapids, MI 49503

\*Contributed equally

Correspondence to:

Brian Haab, PhD

Van Andel Research Institute

333 Bostwick NE

Grand Rapids, MI 49503

616-234-5268

brian.haab@vai.org

Research Supported by:

U01CA152653: National Cancer Institute (Early Detection Research Network

Alliance of Glycobiologists for Cancer Detection, U01CA168896)

National Institute of Allergy and Infectious Diseases (Common Fund Glycoscience Program, R21AI129872).

10 Text Pages, 4 Figures, 0 Tables

## A.1 Abstract

Multimarker fluorescence analysis of tissue specimens offers the opportunity to probe the expression levels and locations of multiple markers in a single sample. Software is needed to fully capitalize on the advantages of this technology for high sensitivity, quantitative, and multiplexed data collection. A major challenge has been the automated identification and quantification of signals. We report software, called SignalFinder, that meets that need. SignalFinder employs a newly-developed algorithm called Segment-Fit Thresholding that demonstrated robust performance for automated signal identification in side-by-side comparisons with several current methods. Two utilities provided with SignalFinder enable downstream analyses. The first allows the quantification and mapping of relationships between an unlimited number of markers through user-defined sequences of AND, OR, and NOT operators. The second produces composite pictures of the signals or colocalization analysis on brightfield H&E images, which is useful for understanding the morphologies and locations of the relevant cells. SignalFinder enables high-throughput, rigorous analyses of whole-slide, multimarker data, and it promises to open new possibilities in various research and clinical applications.

Keywords: Image analysis, immunofluorescence, fluorescence, tissue microarray, automated analysis

## A.2 Introduction

In the analysis of tissue specimens, researchers frequently seek to identify the locations and amounts of specific analytes in the tissue, and then to analyze relationships between the markers and other information. The detection of analytes is usually performed through antibodies that are incubated on the tissue, allowed to bind their targets, and detected by image acquisition. The conventional method of detection is the deposition of colored precipitates produced by the enzymatic conversion of a soluble substrate such as di-aminobenzadine to its insoluble form. The enzyme used for the conversion is often horseradish peroxidase, which is attached to a secondary antibody that localizes to the primary antibody. A brightfield image of the tissue typically shows brown staining on top of cells that are visible through hematoxylin and eosin (H&E) staining. This method has been a workhorse in clinical pathology and research for decades, and it continues to be the primary means of imaging specific proteins in tissue.<sup>220-223</sup>

An increasingly useful and powerful approach for imaging antibody binding is fluorescence. Fluorescence has features that make it preferable to conventional immunohistochemistry in several respects. The signals are very sensitive—especially with continuing improvements in fluorescence microscopes and scanners—and are reproducibly and reliably reflective of analyte concentrations over a broad range. Multiple fluorescence wavelengths can be distinguished from each other with effectively zero crosstalk, which enables the multiplexed detection of multiple analytes in one image, and multiplexing can be greatly expanded through sequential rounds of fluorescence quenching and restaining. Furthermore, fluorescence signals do not obscure the brightfield images of the underlying cells, as can happen with conventional staining. The above features of fluorescence require software for image analysis. A variety of software options currently exist,<sup>224,225</sup> but a particular challenge has been automation—the ability to accurately identify and quantify signals across all images without user intervention or adjustments. Automation is important for increasing throughput and statistical rigor. It is

necessary to remove the potential of user bias and to introduce truly objective analyses of large images and datasets, but it has been difficult to achieve. The largest challenge is the huge variability among images. Some images have much signal, others have little; some have signal-producing features with unforeseen shapes or sizes; and images can have greatly varying backgrounds. In order to automate signal detection, it is necessary to have a fixed basis for determining what is signal that functions across all such characteristics. Basing the signal-detection algorithm on assumed characteristics of the true signals therefore tends to fail for certain images. Preset parameters may function properly for many images, but they eventually require adjustments of settings by the user. Such a requirement limits throughput and potentially introduces bias, and it necessarily brings some level of subjectivity and arbitrariness to the analysis.

We previously introduced an algorithm that does not rely on assumptions about signal characteristics but instead is based on properties of non-signal, or background, regions.<sup>129</sup> It finds non-signal regions through assumptions about the statistical characteristics of background, which are considerably more predictable than those of signals. The algorithm uses the background regions to properly set thresholds for true signals in that image. Thus, the thresholds are tailored precisely to match each image. Variations between images in intensity, amount, shape, or distribution of signal or background are properly accounted for.

In previous work<sup>129,155</sup> we implemented initial versions of software for analyzing multicolor immunofluorescence data and microarray data. The previous software operated well but needed improvements. Because of the computationally-intensive nature of the algorithm, the program was too slow to be applied to large images. Large, high-resolution images are more regularly acquired with the broader availability of whole-slide fluorescence scanners. Also, the output formats are varied across the available fluorescence scanners and microscopes that are commercially available, and we needed to expand the range of image formats that could be

processed.

The SignalFinder software package presented here addresses the above limitations and has significant, new capabilities for immunofluorescence analysis. We achieved major speed advances through code optimization and parallel processing, and we included broad compatibility for all standard types of fluorescence images. In addition, we aimed to provide a full system of analyzing fluorescence images, which includes downstream analyses and visualization. The identification and quantification of signals often is simply the first step; the information must be evaluated in context among multiple signals and among the cells producing the signals. Our system includes two utilities that use the output of the main SignalFinder program. One is for analyzing and quantifying colocalization among distinct signals, and the other is for preparing composites of the brightfield images of the tissue overlaid with the fluorescence signals. This software package is ready for use on Linux, Mac, and Windows operating systems.

In this work, we present the basic capabilities of the package for multicolor fluorescence image analysis and a head-to-head comparison with several other image analysis methods. We compared the automation ability, using only preset parameters, without any user review or adjustment. We furthermore demonstrate the unique capabilities of the software for quantifying and visualizing relationships between distinct markers.

## A.3 Materials and Methods

### A.3.1 Software development and data analysis

We developed and tested the software to implement SFT using MATLAB, supplemented with the image processing and curve fitting toolboxes, Java, and C++. We used Microsoft Excel for analyzing numerical output, GraphPad Prism for the preparation of graphs, and Canvas XIV for the preparation of figures.

### A.3.2 Immunofluorescence data and image processing

The immunofluorescence data had been acquired previously,<sup>155</sup> briefly summarized here. We

performed immunofluorescence on 5  $\mu\text{m}$  thick sections cut from formalin-fixed, paraffin-embedded blocks. We labeled two primary antibodies respectively with Sulfo-Cyanine5 NHS ester (13320, Lumiprobe) and Sulfo-Cyanine3 NHS ester (11320, Lumiprobe) according to the supplier protocol. Each round of immunofluorescence used two different antibodies and nuclear staining with Hoechst 33258. After staining, we scanned the slides using a scanning-fluorescence microscope (Vectra, PerkinElmer). The microscope collected 35 images at each field-of-view, each image at a different emission wavelength. We next quenched the fluorescence using 6%  $\text{H}_2\text{O}_2$  in 250 mM sodium bicarbonate (pH 9.5-10), and performed another round of immunofluorescence using two different antibodies. The subsequent incubations and scanning steps were as described above. The hematoxylin and eosin (H&E) staining followed a standard protocol.

From the 35 images captured for each region, we selected the three that corresponded to the emission maxima of Hoechst 33258, Cy3, and Cy5. For each image, SignalFinder creates a map of the locations of pixels containing signal and computes the percentage of tissue-containing pixels that have signal. To arrive at a final number for each core, we averaged over all images for a core.

## A.4 Results

### A.4.1 Flow of SignalFinderIF processing

The overall analysis system includes the core SignalFinder program as well as two utilities for analyzing the output, ColocFinder and Overlay (Fig. A.1A). The SignalFinder program begins with retrieving the color channels defined by the user, and then analyzing each channel separately in order to find the background and signal pixels. ColocFinder is for analyzing the relationships between the signals from individual channels, and Overlay is for producing composites of the individual or colocalized signals overlaid on the brightfield images.

The package is well suited to analyze tissue microarrays (TMAs) or whole slides with multiple, separate pieces of tissue (Fig. A.1B). TMAs are useful for acquiring data on many tissue

specimens on a single slide, but the amount of information from one experiment can be overwhelming if processed manually or require subsampling already limited tissue. SignalFinder detects tissue cores automatically, or with assistance from the user if necessary, and the image data for each tissue core are analyzed independently. For proper quantification and normalization of the amount of signal, SignalFinder determines the amount of tissue present. This step is important because the image data for a core can include regions with no tissue, causing signal and signal per pixel to be skewed by the blank slide space. Sometimes only a minimal portion of the image contains tissue, for example if the specimen is fragmented or partly washed off. SignalFinder detects regions of non-tissue background, and then it divides the number of signal pixels by the number of tissue pixels to arrive at its final output (Fig. A.1B).

#### A.4.2 Accuracy in automated analysis

The challenge in automated image analysis is being able to handle a wide range of image characteristics accurately, without user review and adjustments of settings. TMA data provide a good test of this capability, since the individual tissue specimens have many, varied characteristics. We used immunofluorescence data from five different TMAs, selecting 11 different 1-mm tissue cores. The acquisition of the images had been performed in earlier work,<sup>155</sup> and it involved the multiplexed detection of 4 different markers plus a nuclear stain on each TMA. The acquisition of the data occurred in two rounds of staining, a method that enables multimarker immunofluorescence (IF) using a limited number of distinct fluorophores.<sup>126</sup> We were particularly interested in a glycan called sTRA that in previous research we identified as a strong serological biomarker of pancreatic cancer.<sup>127</sup> It performed as well as the current best serological biomarker for pancreatic cancer, CA19-9, which also detects a glycan, and it was elevated in about half of the patients with low CA19-9, indicating independent regulation. In addition to data for the two glycans, we also acquired data for the proteins MUC5AC, beta-catenin, vimentin, and E-cadherin (Fig. A.2A). Between the four markers for each TMA and the

two rounds of nuclear staining for 11 different cores, we analyzed 66 different images for this study (Fig. A.2A).

We compared SignalFinder to four commonly-used methods for image analysis of immunofluorescence data. The comparison methods were 1) the ISODATA method<sup>226</sup>; 2) the Huang threshold<sup>227</sup>; 3) the Otsu method<sup>228</sup>; and 4) the Triangle threshold<sup>229,230</sup>.

For each of the five algorithms, we used the settings provided by the software for automated image analysis. SignalFinder is designed to handle multiple images from TMAs or similar data, but we needed to write a custom script to process the images using the other methods. A view of the overall correlations in quantified signal across the 187 images showed that the ISODATA, Huang, and Otsu methods correlated with each other, and that SignalFinder had different results from the rest but was most closely related to the Triangle method (Fig. A.2B). The quantified data were generally higher for the comparison methods relative to SignalFinder, with large differences for some of the images (Fig. A.2C).

For demonstration, we examined selected images. SignalFinder picked out the signal pixels in agreement with the raw fluorescence of each color (Fig. A.2D) for the first three markers of core B2 from TMA69. Each of the comparison methods showed selected locations or entire images that set thresholds too low or that were inconsistent between the colors (e.g. Triangle). Each core image comprises the tiled high-magnification fields from the scanning microscope, with a total of 8-9 fields per core. The results between the fields are inconsistent in some cases for each of the comparison methods, but SignalFinder had improved consistency across the fields. The performance was robust over all cores, as shown for representative selections (Fig. A.2E).

#### A.4.3 Analysis of relationships between markers

A valuable feature of fluorescence in comparison to visible stains is the multiplexing capability. Antibodies can be labeled with different dyes to detect distinct targets, which are routinely used to probe 3 or 4 targets in one run. In addition, using the multi-round method used here,

researchers have acquired data from dozens of markers without evidence of interference or crosstalk between markers.<sup>126,231</sup> As a result of this multiplexing capability, an important use of IF experiments is to detect colocalization of fluorescent signals from distinct probes.<sup>232</sup>

We designed software to quantify exclusive expression as well as colocalization (Fig. A.3A).

The ColocFinder utility allows the user to build up expressions of AND, OR, and NOT between scans, and then quantifies the percentage of pixels that fulfill the expression. The AND operator requires signal pixels to be present in both scans, the OR operator requires pixels to be present in either scan, and the NOT operator requires pixels to be present in the first but not the second scan.

We examined the signals from CA19-9, sTRA, CA19-9 in the absence of sTRA (CA19-9 NOT sTRA), sTRA in the absence of CA19-9 (sTRA NOT CA19-9), and colocalized expression of both CA19-9 and sTRA (CA19-9 AND sTRA) (Fig. A.3B). The program first registers the images to be compared to each other containing the results to be evaluated. It then scans a sliding box of user-defined size (i.e. the colocalization radius or desired proximity to evaluate for two or more markers) across the data and evaluates the pixels in each segment according to the user-defined relationship. If a pixel meets the criteria in a minimum number of the segments, it is counted as positive; otherwise it is negative. The three relationships defined above could be examined relative to another marker such as MUC5AC (Fig. A.3B). The pixel maps and quantification show that most of the MUC5AC is colocalized with CA19-9 in the absence of sTRA, but some is colocalized with both. Thus, complex relationships among multiple markers in their combined and exclusive expression can be visualized and quantified by this system. ColocFinder uses a novel algorithm that does not require complete overlap in pixels, so the various outputs are not mutually-exclusive. The flexibility in finding regions fulfilling the search terms is a good option when markers would be expected to be near one another but not necessarily overlapping. Examples would be two extracellular markers, or a membranous and

extracellular marker, or two markers in cells with unpredictable shapes.

#### A.4.4 Production of composite images

Another advantage of fluorescence is that the signals do not hinder the acquisition of a high-quality brightfield image of the underlying cells. Obtaining a good picture of the underlying cells can be important for determining the types and morphologies of the cells producing certain markers. To facilitate this type of analysis, we developed a utility that uses the output of the signal-finding algorithm or the colocalization analysis to produce images of the fluorescence or colocalization data overlaid on the brightfield image. The program registers the SignalFinder or ColocFinder output to the brightfield image using the nuclei, since the nuclei give the most consistent signal from the Hoechst stain in every round. It extracts the appropriate color range for nuclei in the H&E image, and then registers the output data using the nuclei signal from the scan. The registration can be manually adjusted if necessary. The program then creates images of the output overlaid on the H&E picture, using a color scheme set by the user.

Views of the whole tissue or core can provide information on the locations of features of interest (Fig. A.4A). Images zoomed into specific regions can provide information on the morphologies of cells with particular characteristics. For example, the epithelial layer of a gland expressing both CA19-9 and sTRA and secreting MUC5AC into the lumen has cells with little cytoplasm (Region 1, Fig. A.4B), and the epithelium of another gland producing almost exclusively CA19-9 has more columnar cells (Region 2, Fig. A.4B). The cells expressing primary sTRA, and not CA19-9, form small, ill-defined glandular features (Fig. A.4C). The ability to view side-by-side the original H&E with the composite images can help to identify such features.

#### A.5 Discussion

The increasing availability and quality of whole-slide fluorescence scanning has resulted in increased adoption of this powerful technology. We present here software that meets the demand for automated signal detection and flexible downstream analyses. The method used here allowed quantification of image data from multiple markers and multiple TMAs. Such

analyses would be extremely time-consuming to analyze manually, and the analysis would have been only semi-quantitative.

In the analysis presented, each of the signal finding methods evaluates aspects of the distributions of raw signal intensities with the goal of separating background from signal, with some differences. The ISODATA clustering algorithm uses iterative testing to find the greatest Euclidean distance between signal and background clusters. Huang's fuzzy threshold method steps through thresholds and uses an optimization function across thresholds to determine the true valley between the background and signal peaks. Otsu's method takes the maximum point of intra-class variance and steps through thresholds until it finds maximum variance between the two classes on either side of the threshold. The Triangle method draws a line from the peak of the histogram to the tail of the peak and finds the maximum distance from the line to the curve, which sets the threshold at the inflection point at the end of the primary peak. It then takes anything above that as signal. SignalFinder, in contrast, uses sampling of small regions and the fitting across the regions of relationships between statistical parameters. It first finds the background pixels, and then finds signal pixels based on thresholds derived from the background pixels. It also has a method to disallow spurious "spikes" in the data to be counted as signal pixels. The implication of these differences is that SignalFinder, though more processing intensive, provides more robust compensation for localized background variation in immunohistochemistry image analysis.

The novel component provided by the ColocFinder utility allows explorations of relationships between markers that were not possible with previous software. The currently-available software packages typically quantify colocalization between two markers. For example, a frequently-used method introduced by Manders et al.<sup>233</sup> uses correlations between color channels to calculate average overlap in signals. This information is useful, but additional relationships would be important to probe. Of particular interest is the possibility that the exclusive expression of a particular marker, i.e. the presence of one marker in the absence of

another, could be a marker of phenotype.

Given that certain cell types or tissue phenotypes are identified by the absence and presence of certain markers, researchers should find utility in the quantification of both the exclusive and the concurrent expression of various markers. To further complement this information, the Overlay utility readily reveals which cells meet the various marker expression characteristics defined by the user. This analysis enables cell-morphology analyses to be integrated with the quantitative output of SignalFinder and the ColocFinder utility.

Of additional note, other proprietary software packages have continued to proliferate with imaging systems. Although we chose to benchmark against widely available image analysis algorithms, we recognize the existence of many proprietary software analysis systems tied to imaging platforms. One significant advantage of the SignalFinder Suite is that it is platform and image file type agnostic and is capable of performing the same analysis across all imaging platforms.

We foresee such a system having usefulness for a wide range of research and technological applications, such as in the analysis of immunofluorescence signals from cohorts of patients.<sup>234</sup> In clinical applications, automated image analysis could help to remove inter-operator variability or to pick out rare or subtle features. If the user requires precise and objective quantification, or analysis of signals that are difficult to locate by eye, or the analysis of many data sets, automated quantification is preferable.<sup>224,225</sup>

## A.6 Acknowledgements

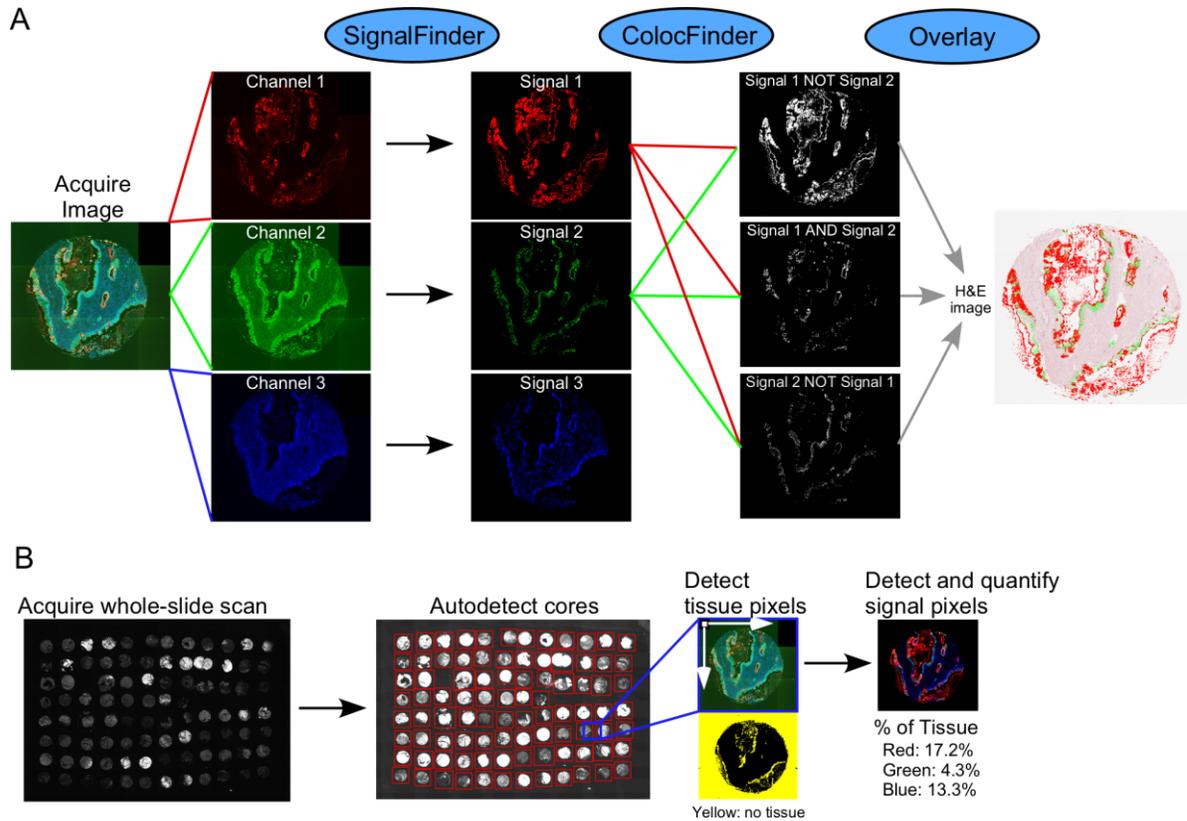
This work was funded by the National Cancer Institute (Early Detection Research Network, U01CA152653; and the Alliance of Glycobiologists for Cancer Detection, U01CA168896) and the National Institute of Allergy and Infectious Diseases (Common Fund Glycoscience Program, R21AI129872).

## A.7 Conflict of Interest Disclosure

The authors declare no competing financial interest.

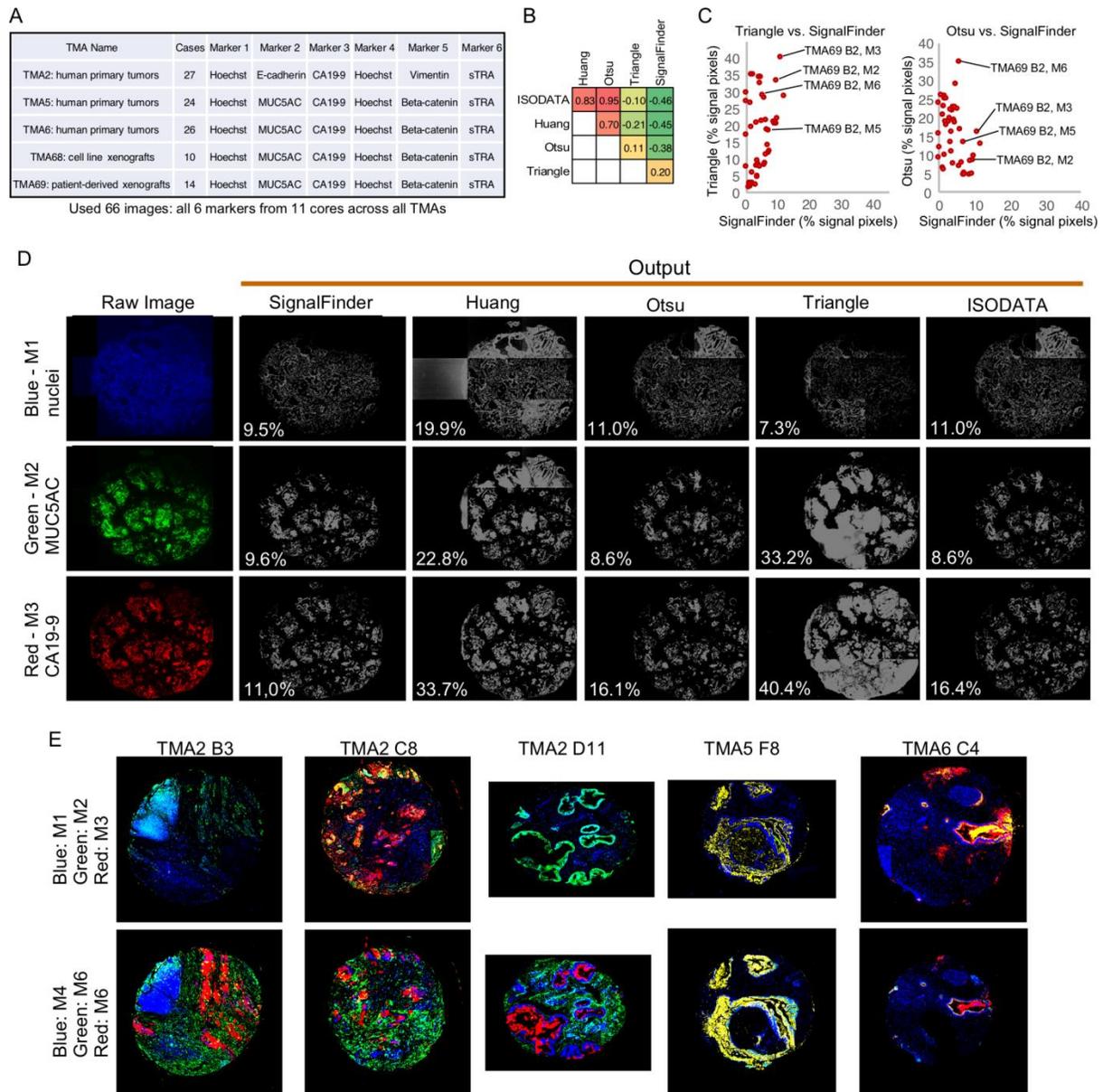
## Methods Figures

Figure A.1 The SignalFinder system.



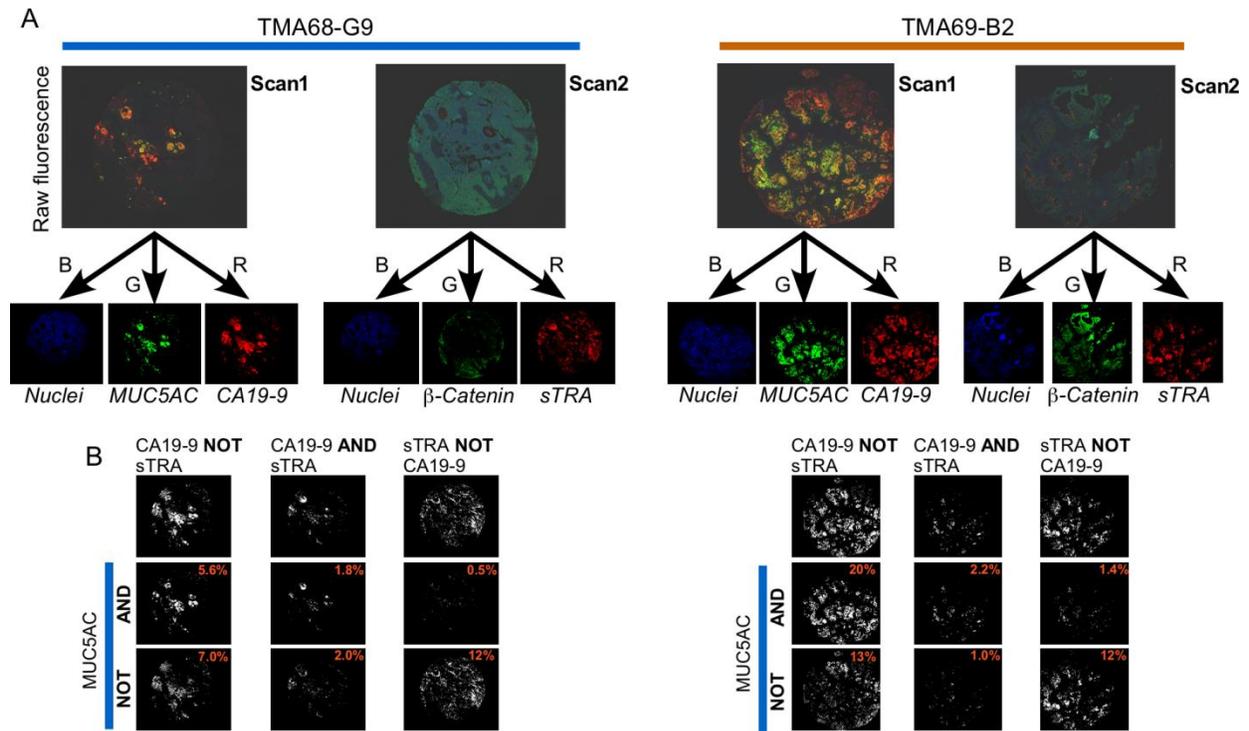
A) The package includes the core SignalFinder program and the utilities Colocfinder and Overlay. SignalFinder separately analyzes individual color channels specified by the user. It identifies and quantifies the signal and produces a map of the signal pixels. ColocFinder identifies regions of the image fulfilling a relationship defined by the user. It quantifies the amount of tissue fulfilling the terms and produces a map of the output pixels. The Overlay program aligns to output from SignalFinder or Coloc finder to a brightfield image and produces a composite, overlaid image using color schemes set by the user.

Figure A.2 Automated image analysis of TMA data.



A) We selected images from 11 cores across the 6 listed TMAs. B) Pearson correlation coefficients for all pairwise comparisons among the five methods. The correlations were calculated across the 66 images. C) The plots present the quantified output from the 11 cores and 4 markers (excluding the Hoechst data). Each point is the %signal from one marker and one core. D) The images are taken from the first three markers for TMA69, core B2. The left column shows the raw data, and the next columns show the output for each method, with the percent of pixels that are positive. E) SignalFinder output from representative cores across the TMAs. The images are the combined signal pixels from the blue, green, and red channels. The overlapping signals have mixed colors. The top images are from the first three markers, and the bottom images are from the second three markers.

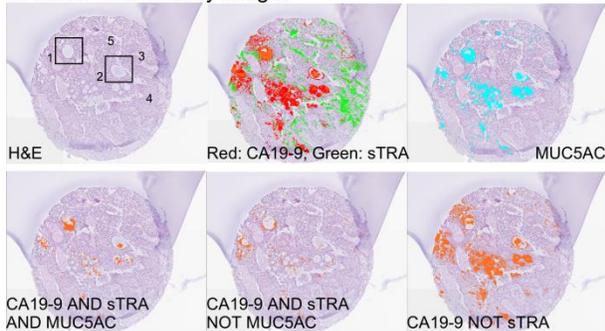
Figure A.3 Exploring relationships between markers.



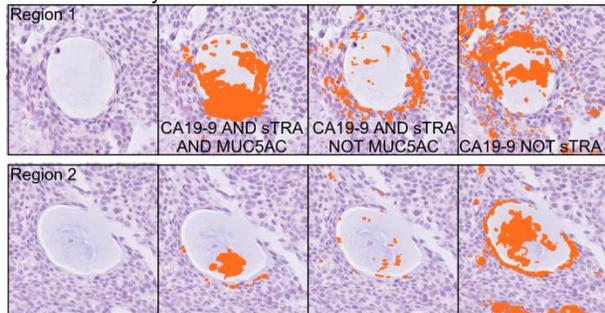
A) Three color channels were acquired in two separate scans, resulting in six markers. The automated signal-finding algorithm identified the signals in the blue, green, and red channels. B) The colocalization utility provides the mapping of user-defined relationships between any number of markers. The example shows the combined and exclusive expression of CA19-9 and sTRA, followed by adding combinations with MUC5AC.

Figure A.4 Composite images.

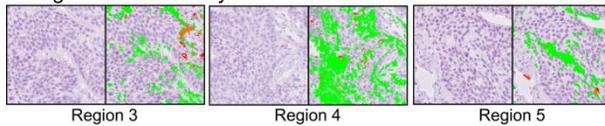
A. Whole-core overlay images



B. Coloc overlays



C. SignalFinder overlays



A) Either the SignalFinder output or the ColocFinder output can be overlaid on the whole-core H&E image. B) Zooms of specific regions provide detailed views of the cells fulfilling various relationships between the markers. C) Detailed views show the unique morphologies of the cells expressing primarily sTRA. The region numbers correspond to those in panel A.

## REFERENCES

## REFERENCES

1. Siegal RL, Miller K, Jemal A. Cancer Statistics 2018. *CA: a cancer journal for clinicians*. 2018;68(1):7-30.
2. Jang JY, Kang MJ, Heo JS, et al. A prospective randomized controlled study comparing outcomes of standard resection and extended resection, including dissection of the nerve plexus and various lymph nodes, in patients with pancreatic head cancer. *Ann Surg*. 2014;259(4):656-664.
3. Cartwright TH, Parisi M, Espirito JL, et al. Clinical Outcomes with First-Line Chemotherapy in a Large Retrospective Study of Patients with Metastatic Pancreatic Cancer Treated in a US Community Oncology Setting. *Drugs - real world outcomes*. 2018;5(3):149-159.
4. Kim S, Signorovitch JE, Yang H, et al. Comparative Effectiveness of nab-Paclitaxel Plus Gemcitabine vs FOLFIRINOX in Metastatic Pancreatic Cancer: A Retrospective Nationwide Chart Review in the United States. *Advances in therapy*. 2018.
5. Yamamoto T, Yagi S, Kinoshita H, et al. Long-term survival after resection of pancreatic cancer: a single-center retrospective analysis. *World journal of gastroenterology*. 2015;21(1):262-268.
6. Raphael BJ, Hruban RH, Aguirre AJ, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer cell*. 2017;32(2).
7. Perera RM, Bardeesy N. Pancreatic Cancer Metabolism: Breaking It Down to Build It Back Up. *Cancer discovery*. 2015;5(12):1247-1261.
8. Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010;467(7319):1114-1117.
9. Noone AM. *SEER Cancer Statistics Review, 1975-2015*. Bethesda, MD: National Cancer Institute; 2018.
10. Guinan P, Bush I, Ray V, Vieth R, Rao R, Bhatti R. The Accuracy of the Rectal Examination in the Diagnosis of Prostate Carcinoma. *New England Journal of Medicine*. 1980;303(9):499-503.
11. Jones D, Friend C, Dreher A, Allgar V, Macleod U. The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: a systematic review. *BMC family practice*. 2018;19(1):79.
12. Walsh AL, Considine SW, Thomas AZ, Lynch TH, Manecksha RP. Digital rectal examination in primary care is important for early detection of prostate cancer: a retrospective cohort analysis study. *The British journal of general practice : the journal of the Royal College of General Practitioners*. 2014;64(629):e783-787.
13. Oesterling JE. Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *The Journal of urology*. 1991;145(5):907-923.
14. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *Jama*. 2005;294(1):66-70.

15. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. *The Journal of urology*. 2013;190(2):419-426.
16. Casanova R, Chuang A, Goepfert AR, et al. *Beckmann and Ling's obstetrics and gynecology*. 2019.
17. History of ACS Recommendations for the Early Detection of Cancer in People Without Symptoms. *American Cancer Society Cancer Prevention and Early Detection Guidelines 2019*; <https://www.cancer.org/health-care-professionals/american-cancer-society-prevention-early-detection-guidelines/overview/chronological-history-of-ac-s-recommendations.html>. Accessed 03/10/19, 2019.
18. Houssami N. Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? *Cancer biology & medicine*. 2017;14(1):1-8.
19. LeMasters T, Sambamoorthi U. A national study of out-of-pocket expenditures for mammography screening. *Journal of women's health (2002)*. 2011;20(12):1775-1783.
20. Kemp Jacobsen K, O'Meara ES, Key D, et al. Comparing sensitivity and specificity of screening mammography in the United States and Denmark. *Int J Cancer*. 2015;137(9):2198-2207.
21. Wong CK, Fedorak RN, Prosser CI, Stewart ME, van Zanten SV, Sadowski DC. The sensitivity and specificity of guaiac and immunochemical fecal occult blood tests for the detection of advanced colonic adenomas and cancer. *International journal of colorectal disease*. 2012;27(12):1657-1664.
22. Issa IA, Nouredine M. Colorectal cancer screening: An updated review of the available options. *World journal of gastroenterology*. 2017;23(28):5086-5096.
23. Force UPST. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement USPSTF Recommendation Statement: Screening for Colorectal Cancer USPSTF Recommendation Statement: Screening for Colorectal Cancer. *Jama*. 2016;315(23):2564-2575.
24. Hoover S, Subramanian S, Tangka FKL, et al. Patients and caregivers costs for colonoscopy-based colorectal cancer screening: Experience of low-income individuals undergoing free colonoscopies. *Evaluation and program planning*. 2017;62:81-86.
25. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*. 2014;371(2):184-188.
26. Sharaf RN, Ladabaum U. Comparative effectiveness and cost-effectiveness of screening colonoscopy vs. sigmoidoscopy and alternative strategies. *Am J Gastroenterol*. 2013;108(1):120-132.
27. Freedman ND, Leitzmann MF, Hollenbeck AR, Schatzkin A, Abnet CC. Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study. *The Lancet. Oncology*. 2008;9(7):649-656.
28. Humphrey L, Deffebach M, Pappas M, et al. Screening for Lung Cancer: Systematic Review to Update the U.S. Preventive

Services Task Force Recommendation. Evidence Synthesis No. 105. AHRQ Publication No. 13-05188-EF-1. In: Quality AfHRa, ed. Rockville, MD2013.

29. Nolen BM, Lokshin AE. Biomarker testing for ovarian cancer: clinical utility of multiplex assays. *Molecular diagnosis & therapy*. 2013;17(3):139-146.
30. Ballehaninna UK, Chamberlain RS. Serum CA 19-9 as a Biomarker for Pancreatic Cancer-A Comprehensive Review. *Indian journal of surgical oncology*. 2011;2(2):88-100.
31. Kosary CL. *SEER Survival Monograph: Cancer Survival Among Adults: US SEER Program, 1988–2001, Patient and Tumor Characteristics*. National Cancer Institute; 2007.
32. Ito T, Igarashi H, Jensen RT. Pancreatic neuroendocrine tumors: clinical features, diagnosis and medical treatment: advances. *Best Pract Res Clin Gastroenterol*. 2012;26(6):737-753.
33. Cancer.Net. Neuroendocrine Tumors of the Pancreas: Introduction. *Neuroendocrine Tumors of the Pancreas* 2017; <https://www.cancer.net/cancer-types/neuroendocrine-tumor-pancreas/introduction>. Accessed 12/3/2018, 2018.
34. Horvath KD, Chabot JA. An aggressive resectional approach to cystic neoplasms of the pancreas. *American Journal of Surgery*. 1999;178(4):269-274.
35. Rodriguez JR, Salvia R, Crippa S, et al. Branch-Duct Intraductal Papillary Mucinous Neoplasms: Observations in 145 Patients Who Underwent Resection. *Gastroenterology*. 2007;133(1):72-79.
36. Hruban RH, Adsay NV. Molecular classification of neoplasms of the pancreas. *Human pathology*. 2009;40(5):612-623.
37. Klimstra DS. Nodular neoplasms of the pancreas. *Modern Pathology*. 2007;20:S94-S112.
38. La Rosa S, Sessa F, Capella C. Acinar Cell Carcinoma of the Pancreas: Overview of Clinicopathologic Features and Insights into the Molecular Pathology. *Front Med*. 2015;2:41.
39. Furukawa T SH, Takeuchi S, et al. Whole exome sequencing reveals recurrent mutations in BRCA2 and FAT genes in acinar cell carcinomas of the pancreas. *Scientific reports*. 2015;5:8829.
40. Bailey JM, DelGiorno KE, Crawford HC. The secret origins and surprising fates of pancreas tumors. *Carcinogenesis*. 2014;35(7):1436-1440.
41. Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology*. 2013;144(6):1252-1261.
42. Aggarwal G, Rabe KG, Petersen GM, Chari ST. New-onset diabetes in pancreatic cancer: A study in the primary care setting. *Pancreatology*. 2012;12(2):156-161.
43. Jacobs EJ, Chanock SJ, Fuchs CS, et al. Family history of cancer and risk of Pancreatic Cancer: A Pooled Analysis from the Pancreatic Cancer Cohort Consortium(PanScan). *International Journal of Cancer*. 2010;127(6):1421-1428.

44. Bosman FT, Carneiro F, Hruban RH, Theise ND. *WHO Classification of Tumors of the Digestive System. Fourth Edition*. Lyon: International Agency for Research on Cancer; 2010.
45. Hruban RH, Goggins M, Parsons J, Kern SE. Progression Model for Pancreatic Cancer. *Clinical Cancer Research*. 2000;6(8):2969-2972.
46. Basturk O, Hong SM, Wood LD, et al. Baltimore Consensus Meeting, A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas. *American Journal of Surgical Pathology*. 2015;39(12):1730-1741.
47. Ferreira RMM, Sancho R, Messal HA, et al. Duct- and Acinar-derived Pancreatic Ductal Adenocarcinomas Show Distinct Tumor Progression and Marker Expression. *Cell Reports*. 2017;21(4):966-978.
48. Dumartin L, Alrawashdeh W, Trabulo SM, et al. ER stress protein AGR2 precedes and is involved in the regulation of pancreatic cancer initiation. *Oncogene*. 2017;36(22):3094-3103.
49. Norris A, Gore A, Balboni A, Young AL, Longnecker DS, Korc M. AGR2 is a SMAD4-suppressible gene that modulates MUC1 levels and promotes the initiation and progression of pancreatic intraepithelial neoplasia. *Oncogene*. 2013;32(33):3867-3876.
50. Kang HJ, Lee JM, Joo I, et al. Assessment of malignant potential in Intraductal Papillary Mucinous Neoplasms of the Pancreas: Comparison between Multidetector CT and MR Imaging with MR Cholangiopancreatography. *Radiology*. 2015;279(1).
51. Molin MD, Matthaei H, Wu J, et al. Clinicopathological correlates of activating GNAS mutations in intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Annals of Surgical Oncology*. 2013;20(12):3802-3808.
52. Wu J, Jiao Y, Dal Molin M, et al. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *PNAS*. 2011;108(52):21188-21193.
53. Poultsides GA, Reddy S, Cameron JL, et al. Histopathologic basis for the favorable survival after resection of intraductal papillary mucinous neoplasm-associated invasive adenocarcinoma of the pancreas. *Annals of Surgery*. 2010;251(3):470-476.
54. Tanaka M. International consensus on the management of intraductal papillary mucinous neoplasm of the pancreas. *Ann Transl Med*. 2015;3(19):286.
55. Elias KM, Tsantoulis P, Tille JC, et al. Primordial germ cells as a potential shared cell of origin for mucinous cystic neoplasms of the pancreas and mucinous ovarian tumors. *Journal of Pathology*. 2018;246(4):459-469 doi: 410.1002/path.5161.
56. Gurzu S, Bara T, Molnar C, et al. epithelial-mesenchymal transition induces aggressivity of mucinous cystic neoplasm of the pancreas with neuroendocrine component: An immunohistochemical study. *Pathology, Research and Practice*. 2018;18:S0344-0338.

57. Kuschner S, Steinle H, Soleiman A, Öfner D, Schneeberger S, Oberhuber G. Intraductal tubulopapillary neoplasm (ITPN) of the pancreas associated with an invasive component: a case report with review of the literature. *World Journal of Surgical Oncology*. 2017;15(1):203.
58. Basturk O, Adsay V, Askan G, al. e. Intraductal Tubulopapillary Neoplasm of the Pancreas: A Clinicopathologic and Immunohistochemical Analysis of 33 Cases. *Am J Surg Pathol*. 2017;41(3):313-325.
59. Patra KC, Bardeesy N, Mizukami Y. Diversity of Precursor Lesions For Pancreatic Cancer: The Genetics and Biology of Intraductal Papillary Mucinous Neoplasm. *Clin Transl Gastroenterol*. 2017;8(4):e86.
60. Iacobuzio-Donahue CA. Genetic evolution of pancreatic cancer: lessons learnt from the pancreatic cancer genome sequencing project. *Gut*. 2012;61(7):1085-1094.
61. Iacobuzio-Donahue CA, Fu B, Yachida S, et al. DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2009;27(11):1806-1813.
62. Blackford A, Serrano OK, Wolfgang CL, al. e. SMAD4 gene mutations are associated with poor prognosis in pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(14):4674-4679.
63. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531(7592):47-52.
64. Wang F, Xia X, Yang C, et al. SMAD4 Gene Mutation Renders Pancreatic Cancer Resistant to Radiotherapy through Promotion of Autophagy. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2018;24(13):3176-3185.
65. Zeng G, Germinaro M, Micsenyi A, et al. Aberrant Wnt/ $\beta$ -Catenin Signaling in Pancreatic Adenocarcinoma. *Neoplasia*. 2006;8(4):279-289.
66. Valent P, Bonnet D, De Maria R, et al. Cancer stem cell definitions and terminology: the devil is in the details. *Nature reviews. Cancer*. 2012;12(11):767-775.
67. Hermann PC, Huber SL, Herrler T, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell*. 2017;1(3):313-323.
68. Bailey JM, Alsina J, Rasheed Z, et al. DCLK1 Marks a Morphologically Distinct Subpopulation of Cells With Stem Cell Properties in Preinvasive Pancreatic Cancer. *Gastroenterology*. 2014;146(1):245-256.
69. Durko L, Wlodarski W, Stasikowska-Kanicka O, al e. Expression and Clinical Significance of Cancer Stem Cell Markers CD24, CD44, and CD133 in Pancreatic Ductal Adenocarcinoma and Chronic Pancreatitis. *Dis Markers*. 2017;2017:3276806.
70. Immervoll H, Hoem D, Steffensen OJ, Miletic H, Molven A. Visualization of CD44 and CD133 in normal pancreas and pancreatic ductal adenocarcinomas: non-overlapping membrane

- expression in cell populations positive for both markers. *J Histochem Cytochem.* 2011;59(44):441-455.
71. Kaur S, Kumar S, Momi N, Sasson AR, Batra SK. Mucins in pancreatic cancer and its microenvironment. *Nat Rev Gastroenterol Hepatol.* 2013;10(10):607-620.
  72. Kleeff J, Ishiwata T, Kumbasar AH, et al. The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *The Journal of clinical investigation.* 1998;102(9):1662-1673.
  73. Oshio G, Ogawa K, Kudo H, et al. Immunohistochemical studies on the localization of cancer associated antigens DU-PAN-2 and CA19-9 in carcinomas of the digestive tract. *J Gastroenterol Hepatol.* 1990;5(1):25-31.
  74. Yue T, Goldstein IJ, Hollingsworth MA, Kaul K, Brand RE, Haab BB. The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Molecular & cellular proteomics : MCP.* 2009;8(7):1697-1670.
  75. Knelson EH, Nee JC, Blobe GC. Heparan sulfate signaling in cancer. *Trends in biochemical sciences.* 2014;39(6):277-288.
  76. Raphael BJ, Hruban RH, Aguirre AJ, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer cell.* 2017;32(2):185-203.e113.
  77. Wood LD, Hruban RH. Pathology and molecular genetics of pancreatic neoplasms. *Cancer journal (Sudbury, Mass.).* 2012;18(6):492-501.
  78. Rasheed ZA, Matsui W, Maitra A. Pathology of pancreatic stroma in PDAC. *Pancreatic Cancer and Tumor Microenvironment.* Trivandrum (India): Transworld Research Network; 2012:Chapter 1.
  79. Laklai H, Miroshnikova YA, Pickup MW, al. e. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. *Nature medicine.* 2016;22(5):497-505.
  80. Northey JJ, L P, Weaver VM. Tissue Force Programs Cell Fate and Tumor Aggression. *Cancer discovery.* 2017;7(11).
  81. Fu Y, Liu S, Zeng S, Shen H. The critical roles of activated stellate cells-mediated paracrine signaling, metabolism and onco-immunology in pancreatic ductal adenocarcinoma. *Molecular cancer.* 2018;17(1):62.
  82. Ren B, Cui M, Yang G, et al. Tumor microenvironment participates in metastasis of pancreatic cancer. *Molecular cancer.* 2018;17(1):108.
  83. Santi A, Kugeratski FG, Zanivan S. Cancer Associated Fibroblasts: The Architects of Stroma Remodeling. *Proteomics.* 2018;18:1700167.

84. Borgoni S, Iannello A, Cutrupi S, et al. Depletion of tumor-associated macrophages switches the epigenetic profile of pancreatic cancer infiltrating T cells and restores their anti-tumor phenotype. *Oncol Immunology*. 2018;7(2).
85. Shiga K, Hara M, Nagasaki T, Sato T, Takahashi H, Takeyama H. Cancer-Associated Fibroblasts: Their Characteristics and Their Roles in Tumor Growth. *Cancers*. 2015;7(4):2443-2458.
86. Wang X, Zhang W, Sun X, Lin Y, Chen W. Cancer-associated fibroblasts induce epithelial-mesenchymal transition through secreted cytokines in endometrial cancer cells. *Oncol Lett*. 2018;15(4):5694-5702.
87. Haerberle L, Steiger K, Schlitter AM, et al. Stromal heterogeneity in pancreatic cancer and chronic pancreatitis. *Pancreatology*. 2018;18(5):536-549.
88. Takahashi D, Kojima M, Suzuki T, et al. Profiling the Tumour Immune Microenvironment in Pancreatic Neuroendocrine Neoplasms with Multispectral Imaging Indicates Distinct Subpopulation Characteristics Concordant with WHO 2017 Classification. *Scientific reports*. 2018;8:13166.
89. Waddell N, Pajic M, Patch AM, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. 2015;518(7540):495-501.
90. Grover S, Syngal S. Hereditary pancreatic cancer. *Gastroenterology*. 2010;139(4):1076-1080, 1080.e1071-1072.
91. Fauquette V, Perrais M, Cerulis S, et al. The antagonistic regulation of human MUC4 and ErbB-2 genes by the Ets protein PEA3 in pancreatic cancer cells: implications for the proliferation/differentiation balance in the cells. *Biochem J*. 2005;386(Pt 1):35-45.
92. Garcia-Carracedo D, Chen ZM, Qiu W, et al. PIK3CA mutations in mucinous cystic neoplasms of the pancreas. *Pancreas*. 2014;43(2):245-249.
93. Collisson EA, Collisson EA, Sadanandam A, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine*. 2011;17(4):500-503.
94. Moffitt RA, Marayati R, Flate EL, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature genetics*. 2015;47(10):1168-1178.
95. McCormick KA, Coveler AL, Rossi GR, Vahanian NN, Link C, Chiorean EG. Pancreatic cancer: Update on immunotherapies and algenpantucel-L. *Human vaccines & immunotherapeutics*. 2016;12(3):563-575.
96. Cote GA, Gore AJ, McElyea SD, et al. A Pilot Study to Develop a Diagnostic Test for Pancreatic Ductal Adenocarcinoma Based on Differential Expression of Select miRNA in Plasma and Bile. *The American Journal of Gastroenterology*. 2014;109(12):1942-1952.
97. Hu G-y, Tao F, Wang W, Ji K-w. Prognostic value of microRNA-21 in pancreatic ductal adenocarcinoma: a meta-analysis. *World Journal of Surgical Oncology*. 2016;14(1):82-82.

98. Kudo T. Molecular Genetic Analysis of the Human Lewis Histo-blood Group System. *Journal of Biological Chemistry*. 1996;271(16):9830 –9837.
99. Ballehaninna UK, Chamberlain RS. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. *Journal of gastrointestinal oncology*. 2012;3(2):105-119.
100. Passerini R, Cassatella MC, Boveri S, et al. The Pitfalls of CA19-9 Routine Testing and Comparison of Two Automated Immunoassays in a Reference Oncology Center. *American Journal of Clinical Pathology*. 2012;138(2):281-287.
101. Guo M, Luo G, Lu R, et al. Distribution of Lewis and Secretor polymorphisms and corresponding CA19-9 antigen expression in a Chinese population. *FEBS open bio*. 2017;7(11):1660-1671.
102. Luo G, Liu C, Guo M, et al. Potential Biomarkers in Lewis Negative Patients With Pancreatic Cancer. *Ann Surg*. 2017;265(4):800-805.
103. Vestergaard EM, Hein HO, Meyer H, et al. Reference values and biological variation for tumor marker CA 19-9 in serum for different Lewis and secretor genotypes and evaluation of secretor and Lewis genotyping in a Caucasian population. *Clinical chemistry*. 1999;45(1):54-61.
104. Tang H, Partyka K, Hsueh P, et al. Glycans Related to the CA19-9 Antigen Are Increased in Distinct Subsets of Pancreatic Cancers and Improve Diagnostic Accuracy Over CA19-9. *Cellular and molecular gastroenterology and hepatology*. 2016;2(2):201.
105. Cummings RD, Stanley P. Chapter 13 Structures Common to Different Glycans. *Essentials of Glycobiology 2nd Ed.* . NY: Cold Spring Harbor Laboratory Press; 2009.
106. Andrews PW, Banting G, Damjanov I, Arnaud D, Avner P. Three Monoclonal Antibodies Defining Distinct Differentiation Antigens Associated with Different High Molecular Weight Polypeptides on the Surface of Human Embryonal Carcinoma Cells. *Hybridoma*. 1984;3(4).
107. Bouhassira E. *The Sage Encyclopedia of Stem Cell Research (Vols. 1-3)*. Thousand Oaks, CA: SAGE Publications Ltd. ; 2015.
108. Natunen S, Satomaa T, Pitkanen V, et al. The binding specificity of the marker antibodies Tra-1-60 and Tra-1-81 reveals a novel pluripotency-associated type 1 lactosamine epitope. *Glycobiology*. 2011;21(9):1125-1130.
109. Haab B. Primary Screen: primscreen\_6035. *Consortium for Functional Glycomics*. 2013. [http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen\\_6035](http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen_6035).
110. Louis N. Primary Screen: primscreen\_3257 *Consortium for Functional Glycomics*. 2010. [http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen\\_3257](http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen_3257).
111. Mahal LK. Primary Screen: primscreen\_5465. *Consortium for Functional Glycomics*. 2011. [http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen\\_5465](http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen_5465).

112. Nakata B, Wang YQ, Yashiro M, et al. Negative hMSH2 protein expression in pancreatic carcinoma may predict a better prognosis of patients. *Oncology reports*. 2003;10(4):997-1000.
113. Wild AT, Dholakia AS, Fan KY, et al. Efficacy of platinum chemotherapy agents in the adjuvant setting for adenosquamous carcinoma of the pancreas. *Journal of gastrointestinal oncology*. 2015;6(2):115-125.
114. Adsay NV, Pierson C, Sarkar F, et al. Colloid (mucinous noncystic) carcinoma of the pancreas. *The American journal of surgical pathology*. 2001;25(1):26-42.
115. Krasinskas AM, Moser AJ, Saka B, Adsay NV, Chiosea SI. KRAS mutant allele-specific imbalance is associated with worse prognosis in pancreatic cancer and progression to undifferentiated carcinoma of the pancreas. *Mod Pathol*. 2013;26(10):1346-1354.
116. Muraki T, Reid MD, Basturk O, et al. Undifferentiated Carcinoma With Osteoclastic Giant Cells of the Pancreas: Clinicopathologic Analysis of 38 Cases Highlights a More Protracted Clinical Course Than Currently Appreciated. *The American journal of surgical pathology*. 2016;40(9):1203-1216.
117. Villarreal MC, Rajeshkumar NV, Garrido-Laguna I, et al. Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. *Molecular cancer therapeutics*. 2011;10(1):3-8.
118. Golan T, Kanji ZS, Epelbaum R, et al. Overall survival and clinical characteristics of pancreatic cancer in BRCA mutation carriers. *British journal of cancer*. 2014;111(6):1132-1138.
119. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine*. 2015;372(26):2509-2520.
120. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine*. 2011;4:500-503.
121. Waddell N, Pajic M, Patch AM, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. 2015;518(7540):495-501.
122. Li C, Heidt DG, Dalerba P, et al. Identification of pancreatic cancer stem cells. *Cancer research*. 2007;67(3):1030-1037.
123. Rasheed ZA, Yang J, Wang Q, et al. Prognostic Significance of Tumorigenic Cells With Mesenchymal Features in Pancreatic Adenocarcinoma. *Journal of the National Cancer Institute*. 2010.
124. Rhim AD, Mirek ET, Aiello NM, et al. EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell*. 2012;148(1-2):349-361.
125. Yu M, Ting DT, Stott SL, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature*. 2012;487(7408):510-513.
126. Gerdes MJ, Sevinsky CJ, Sood A, et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(29):11982-11987.

127. Tang H, Partyka K, Hsueh P, et al. Glycans related to the CA19-9 antigen are elevated in distinct subsets of pancreatic cancers and improve diagnostic accuracy over CA19-9. *Cell Mol Gastroenterol Hepatol*. 2016;2(2):201-221 e215.
128. Andrews PW, Banting G, Damjanov I, Arnaud D, Avner P. Three monoclonal antibodies defining distinct differentiation antigens associated with different high molecular weight polypeptides on the surface of human embryonal carcinoma cells. *Hybridoma*. 1984;3(4):347-361.
129. Ensink E, Sinha J, Sinha A, et al. Segment and fit thresholding: a new method for image analysis applied to microarray and immunofluorescence data. *Analytical chemistry*. 2015;87(19):9715-9721.
130. Dursun N, Feng J, Basturk O, Bandyopadhyay S, Cheng JD, Adsay VN. Vacuolated cell pattern of pancreatobiliary adenocarcinoma: a clinicopathological analysis of 24 cases of a poorly recognized distinctive morphologic variant important in the differential diagnosis. *Virchows Arch*. 2010;457(6):643-649.
131. Adsay V, Logani S, Sarkar F, Crissman J, Vaitkevicius V. Foamy gland pattern of pancreatic ductal adenocarcinoma: a deceptively benign-appearing variant. *The American journal of surgical pathology*. 2000;24(4):493-504.
132. Fallon BP, Curnutte B, Maupin KA, et al. The Marker State Space (MSS) Method for Classifying Clinical Samples. *PLoS ONE*. 2013;8(6):e65905.
133. Jiang JH, Liu C, Cheng H, et al. Epithelial-mesenchymal transition in pancreatic cancer: Is it a clinically significant factor? *Biochimica et biophysica acta*. 2015;1855(1):43-49.
134. McDonald OG, Maitra A, Hruban RH. Human correlates of provocative questions in pancreatic pathology. *Advances in anatomic pathology*. 2012;19(6):351-362.
135. Winter JM, Tang LH, Klimstra DS, et al. A novel survival-based tissue microarray of pancreatic cancer validates MUC1 and mesothelin as biomarkers. *PLoS ONE*. 2012;7(7):e40157.
136. Winter JM, Yeo CJ, Brody JR. Diagnostic, prognostic, and predictive biomarkers in pancreatic cancer. *Journal of surgical oncology*. 2013;107(1):15-22.
137. Ansari D, Rosendahl A, Elebro J, Andersson R. Systematic review of immunohistochemical biomarkers to identify prognostic subgroups of patients with pancreatic cancer. *Br J Surg*. 2011.
138. McCarthy NC, Albrechtsen MT, Kerr MA. Characterization of a human granulocyte differentiation antigen (CDw15) commonly recognized by monoclonal antibodies. *Bioscience reports*. 1985;5(10-11):933-941.
139. Shablott MJ, Axelman J, Wang S, et al. Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(23):13726-13731.
140. Mazzetti S, Frigerio S, Gelati M, Salmaggi A, Vitellaro-Zuccarello L. Lycopersicon esculentum lectin: an effective and versatile endothelial marker of normal and tumoral blood vessels in the central nervous system. *European journal of histochemistry : EJH*. 2004;48(4):423-428.

141. Tempero MA, Uchida E, Takasaki H, Burnett DA, Steplewski Z, Pour PM. Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. *Cancer research*. 1987;47(20):5501-5503.
142. Pour PM, Tempero MM, Takasaki H, et al. Expression of blood group-related antigens ABH, Lewis A, Lewis B, Lewis X, Lewis Y, and CA 19-9 in pancreatic cancer cells in comparison with the patient's blood group type. *Cancer research*. 1988;48(19):5422-5426.
143. Nilsson O, Lindholm L, Holmgren J, Svennerholm L. Monoclonal antibodies raised against NeuAc alpha 2-6neolactotetraosylceramide detect carcinoma-associated gangliosides. *Biochimica et biophysica acta*. 1985;835(3):577-583.
144. Fredman P, von Holst H, Collins VP, Granholm L, Svennerholm L. Sialyllactotetraosylceramide, a ganglioside marker for human malignant gliomas. *Journal of neurochemistry*. 1988;50(3):912-919.
145. Hansson GC, Zopf D. Biosynthesis of the cancer-associated sialyl-Lea antigen. *The Journal of biological chemistry*. 1985;260(16):9388-9392.
146. McEver RP. Selectin-carbohydrate interactions during inflammation and metastasis. *Glycoconjugate journal*. 1997;14(5):585-591.
147. Bhat R, Belardi B, Mori H, et al. Nuclear repartitioning of galectin-1 by an extracellular glycan switch regulates mammary morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(33):E4820-4827.
148. Monsma DJ, Monks NR, Cherba DM, et al. Genomic characterization of explant tumorgraft models derived from fresh patient tumor tissue. *J Transl Med*. 2012;10:125.
149. Goonetilleke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol*. 2007;33(3):266-270.
150. Steinberg W. The clinical utility of the CA 19-9 tumor-associated antigen. *The American journal of gastroenterology*. 1990;85(4):350-355.
151. Malesci A, Montorsi M, Mariani A, et al. Clinical utility of the serum CA 19-9 test for diagnosing pancreatic carcinoma in symptomatic patients: a prospective study. *Pancreas*. 1992;7(4):497-502.
152. Haab BB, Huang Y, Balasenthil S, et al. Definitive Characterization of CA 19-9 in Resectable Pancreatic Cancer Using a Reference Set of Serum and Plasma Specimens. *PLoS One*. 2015;10(10):e0139049.
153. Tang H, Singh S, Partyka K, et al. Glycan motif profiling reveals plasma sialyl-Lewis X elevations in pancreatic cancers that are negative for CA 19-9. *Mol Cell Proteomics*. 2015;14(5):1323-1333.
154. Singh S, Pal K, Yadav J, et al. Upregulation of glycans containing 3' fucose in a subset of pancreatic cancers uncovered using fusion-tagged lectins. *Journal of proteome research*. 2015;14(6):2594-2605.

155. Barnett D, Liu Y, Partyka K, et al. The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells. *Scientific reports*. 2017;7(1):4020.
156. Lennon AM, Wolfgang CL, Canto MI, et al. The Early Detection of Pancreatic Cancer: What Will It Take to Diagnose and Treat Curable Pancreatic Neoplasia? *Cancer research*. 2014.
157. Kelly KA, Hollingsworth MA, Brand RE, et al. Advances in Biomedical Imaging, Bioengineering, and Related Technologies for the Development of Biomarkers of Pancreatic Disease: Summary of a National Institute of Diabetes and Digestive and Kidney Diseases and National Institute of Biomedical Imaging and Bioengineering Workshop. *Pancreas*. 2015;44(8):1185-1194.
158. Young MR, Wagner PD, Ghosh S, et al. Validation of Biomarkers for Early Detection of Pancreatic Cancer: Summary of The Alliance of Pancreatic Cancer Consortia for Biomarkers for Early Detection Workshop. *Pancreas*. 2018;47(2):135-141.
159. Chen S, LaRoche T, Hamelinck D, et al. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nature methods*. 2007;4(5):437-444.
160. Yue T, Goldstein IJ, Hollingsworth MA, Kaul K, Brand RE, Haab BB. The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics*. 2009;8(7):1697-1707.
161. Yue T, Maupin KA, Fallon B, et al. Enhanced discrimination of malignant from benign pancreatic disease by measuring the CA 19-9 antigen on specific protein carriers. *PLoS ONE*. 2011;6(12):e29180.
162. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. 1st ed. ed. Boca Raton, FL: CRC Press; 1994.
163. Herlyn M, Steplewski Z, Herlyn D, Koprowski H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proceedings of the National Academy of Sciences of the United States of America*. 1979;76(3):1438-1442.
164. Magnani JL, Nilsson B, Brockhaus M, et al. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. *The Journal of biological chemistry*. 1982;257(23):14365-14369.
165. Magnani JL, Brockhaus M, Smith DF, et al. A monosialoganglioside is a monoclonal antibody-defined antigen of colon carcinoma. *Science (New York, N.Y.)*. 1981;212(4490):55-56.
166. Yue T, Partyka K, Maupin KA, et al. Identification of blood-protein carriers of the CA 19-9 antigen and characterization of prevalence in pancreatic diseases. *Proteomics*. 2011;11(18):3665-3674.
167. Sah RP, Nagpal SJ, Mukhopadhyay D, Chari ST. New insights into pancreatic cancer-induced paraneoplastic diabetes. *Nat Rev Gastroenterol Hepatol*. 2013;10(7):423-433.
168. Chari ST, Leibson CL, Rabe KG, Ransom J, de Andrade M, Petersen GM. Probability of pancreatic cancer following diabetes: a population-based study. *Gastroenterology*. 2005;129(2):504-511.

169. Kim J, Bamlet WR, Oberg AL, et al. Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. *Sci Transl Med.* 2017;9(398).
170. Honda K, Kobayashi M, Okusaka T, et al. Plasma biomarker for detection of early stage pancreatic cancer and risk factors for pancreatic malignancy using antibodies for apolipoprotein-All isoforms. *Scientific reports.* 2015;5:15921.
171. Capello M, Bantis LE, Scelo G, et al. Sequential Validation of Blood-Based Protein Biomarker Candidates for Early-Stage Pancreatic Cancer. *Journal of the National Cancer Institute.* 2017;109(4).
172. Balasenthil S, Huang Y, Liu S, et al. A Plasma Biomarker Panel to Identify Surgically Resectable Early-Stage Pancreatic Cancer. *Journal of the National Cancer Institute.* 2017;109(8).
173. Cohen JD, Javed AA, Thoburn C, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proceedings of the National Academy of Sciences of the United States of America.* 2017;114(38):10202-10207.
174. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science (New York, N.Y.)* 2018;359(6378):926-930.
175. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *Journal of the National Cancer Institute.* 2008;100(20):1432-1438.
176. Sullivan Pepe M, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute.* 2001;93(14):1054-1061.
177. Balmana M, Sarrats A, Llop E, et al. Identification of potential pancreatic cancer serum markers: Increased sialyl-Lewis X on ceruloplasmin. *Clinica chimica acta; international journal of clinical chemistry.* 2015;442C:56-62.
178. Metzgar RS, Gaillard MT, Levine SJ, Tuck FL, Bossen EH, Borowitz MJ. Antigens of human pancreatic adenocarcinoma cells defined by murine monoclonal antibodies. *Cancer research.* 1982;42(2):601-608.
179. Kawa S, Tokoo M, Oguchi H, et al. Epitope analysis of SPan-1 and DUPAN-2 using synthesized glycoconjugates sialyllact-N-fucopentaose II and sialyllact-N-tetraose. *Pancreas.* 1994;9(6):692-697.
180. Takasaki H, Uchida E, Tempero MA, Burnett DA, Metzgar RS, Pour PM. Correlative study on expression of CA 19-9 and DU-PAN-2 in tumor tissue and in serum of pancreatic cancer patients. *Cancer research.* 1988;48(6):1435-1438.
181. Partyka K, Maupin KA, Brand RE, Haab BB. Diverse monoclonal antibodies against the CA 19-9 antigen show variation in binding specificity with consequences for clinical interpretation. *Proteomics.* 2012;12:2212-2220.

182. Haglund C, Lindgren J, Roberts PJ, Nordling S. Gastrointestinal cancer-associated antigen CA 19-9 in histological specimens of pancreatic tumours and pancreatitis. *British journal of cancer*. 1986;53(2):189-195.
183. Kalthoff H, Kreiker C, Schmiegel WH, Greten H, Thiele HG. Characterization of CA 19-9 bearing mucins as physiological exocrine pancreatic secretion products. *Cancer research*. 1986;46(7):3605-3607.
184. Powers TW, Jones EE, Betesh LR, et al. Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked glycan expression in tissues. *Analytical chemistry*. 2013;85(20):9799-9806.
185. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: a cancer journal for clinicians*. 2018;68(1):7-30.
186. Taddei ML, Giannoni E, Comito G, Chiarugi P. Microenvironment and tumor cell plasticity: an easy way out. *Cancer letters*. 2013;341(1):80-96.
187. Olive KP, Jacobetz MA, Davidson CJ, et al. Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science (New York, N.Y.)*. 2009;324(5933):1457-1461.
188. Rhim AD, Oberstein PE, Thomas DH, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer cell*. 2014;25(6):735-747.
189. Gu D, Schlotman KE, Xie J. Deciphering the role of hedgehog signaling in pancreatic cancer. *Journal of biomedical research*. 2016;30(5):353-360.
190. Loos M, Giese NA, Kleeff J, et al. Clinical significance and regulation of the costimulatory molecule B7-H1 in pancreatic cancer. *Cancer letters*. 2008;268(1):98-109.
191. Kim MS, Zhong Y, Yachida S, et al. Heterogeneity of pancreatic cancer metastases in a single patient revealed by quantitative proteomics. *Molecular & cellular proteomics : MCP*. 2014;13(11):2803-2811.
192. Lin WC, Rajbhandari N, Liu C, et al. Dormant cancer cells contribute to residual disease in a model of reversible pancreatic cancer. *Cancer research*. 2013;73(6):1821-1830.
193. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine*. 2011;17(4):500-503.
194. Staal B, Liu Y, Barnett D, et al. The sTRA Plasma Biomarker: Blinded Validation of Improved Accuracy over CA19-9 in Pancreatic Cancer Diagnosis. 2018. *Clinical Cancer Research (Accepted for Publication)*. 2018.
195. Wu YM, Nowack DD, Omenn GS, Haab BB. Mucin glycosylation is altered by pro-inflammatory signaling in pancreatic-cancer cells. *Journal of proteome research*. 2009;8(4):1876-1886.

196. Barnett D, Hall J, Haab BB. Automated identification and quantification of signals in multichannel immunofluorescence images: the SignalFinder platform. (*Manuscript Submitted for Publication*). 2018.
197. Partyka K, Maupin KA, Brand RE, Haab BB. Diverse monoclonal antibodies against the CA 19-9 antigen show variation in binding specificity with consequences for clinical interpretation. *Proteomics*. 2012;12(13):2212-2220.
198. Xiao Z, Luo G, Liu C, et al. Molecular mechanism underlying lymphatic metastasis in pancreatic cancer. *BioMed research international*. 2014;2014:925845.
199. Le Large TYS, Bijlsma MF, Kazemier G, van Laarhoven HWM, Giovannetti E, Jimenez CR. Key biological processes driving metastatic spread of pancreatic cancer as identified by multi-omics studies. *Seminars in cancer biology*. 2017;44:153-169.
200. Trinchera M, Aronica A, Dall'Olio F. Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology*. 2017;6(1).
201. Atlas P. "SELL". *The Human Protein Atlas 2018*; <https://www.proteinatlas.org/ENSG00000188404-SELL/tissue>. Accessed November 30, 2018.
202. Kannagi R. Carbohydrate-mediated cell adhesion involved in hematogenous metastasis of cancer. *Glycoconjugate journal*. 1997;14(5):577-584.
203. Varki A. Selectin ligands: will the real ones please stand up? *The Journal of clinical investigation*. 1997;99(2):158-162.
204. Hosch SB, Knoefel WT, Metz S, et al. Early lymphatic tumor cell dissemination in pancreatic cancer: frequency and prognostic significance. *Pancreas*. 1997;15(2):154-159.
205. Furukawa H, Okada S, Saisho H, et al. Clinicopathologic features of small pancreatic adenocarcinoma. A collective study. *Cancer*. 1996;78(5):986-990.
206. Ando N, Nakao A, Nomoto S, et al. Detection of mutant K-ras in dissected paraaortic lymph nodes of patients with pancreatic adenocarcinoma. *Pancreas*. 1997;15(4):374-378.
207. von Haehling S, Anker MS, Anker SD. Prevalence and clinical impact of cachexia in chronic illness in Europe, USA, and Japan: facts and numbers update 2016. *Journal of cachexia, sarcopenia and muscle*. 2016;7(5):507-509.
208. von Haehling S, Anker SD. Cachexia as a major underestimated and unmet medical need: facts and numbers. *Journal of cachexia, sarcopenia and muscle*. 2010;1(1):1-5.
209. Strobel O, Rosow DE, Rakhlin EY, et al. Pancreatic duct glands are distinct ductal compartments that react to chronic injury and mediate Shh-induced metaplasia. *Gastroenterology*. 2010;138(3):1166-1177.
210. Xing L, Shi Q, Zheng K, et al. Ultrasound-Mediated Microbubble Destruction (UMMD) Facilitates the Delivery of CA19-9 Targeted and Paclitaxel Loaded mPEG-PLGA-PLL Nanoparticles in Pancreatic Cancer. *Theranostics*. 2016;6(10):1573-1587.

211. Yoshida M, Takimoto R, Murase K, et al. Targeting anticancer drug delivery to pancreatic cancer cells using a fucose-bound nanoparticle approach. *PLoS one*. 2012;7(7):e39545.
212. Ho JJ, Siddiki B, Kim YS. Association of sialyl-Lewis(a) and sialyl-Lewis(x) with MUC-1 apomucin in a pancreatic cancer cell line. *Cancer research*. 1995;55(16):3659-3663.
213. Girgis MD, Olafsen T, Kenanova V, McCabe KE, Wu AM, Tomlinson JS. CA19-9 as a Potential Target for Radiolabeled Antibody-Based Positron Emission Tomography of Pancreas Cancer. *International journal of molecular imaging*. 2011;2011:834515.
214. Costa-Silva B, Aiello NM, Ocean AJ, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature cell biology*. 2015;17(6):816-826.
215. Ravn V, Dabelsteen E. Tissue distribution of histo-blood group antigens. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. 2000;108(1):1-28.
216. Ito N, Hirota T. Histochemical and cytochemical localization of blood group antigens. *Progress in histochemistry and cytochemistry*. 1992;25(2):1-85.
217. Zhang W, Zhu ZY. Structural modification of H histo-blood group antigen. *Blood transfusion = Trasfusione del sangue*. 2015;13(1):143-149.
218. Sawada R, Sun SM, Wu X, et al. Human monoclonal antibodies to sialyl-Lewis (CA19.9) with potent CDC, ADCC, and antitumor activity. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011;17(5):1024-1032.
219. Tiriac H, Belleau P, Engle DD, et al. Organoid Profiling Identifies Common Responders to Chemotherapy in Pancreatic Cancer. *Cancer discovery*. 2018;8(9):1112-1129.
220. McCourt CM, Boyle D, James J, Salto-Tellez M. Immunohistochemistry in the era of personalised medicine. *Journal of clinical pathology*. 2013;66(1):58-61.
221. Hamilton PW, Bankhead P, Wang Y, et al. Digital pathology and image analysis in tissue biomarker research. *Methods*. 2014;70(1):59-73.
222. Salto-Tellez M. Diagnostic Molecular Cytopathology - a further decade of progress. *Cytopathology : official journal of the British Society for Clinical Cytology*. 2015;26(5):269-270.
223. Jones JL, Oien KA, Lee JL, Salto-Tellez M. Morphomolecular pathology: setting the framework for a new generation of pathologists. *British journal of cancer*. 2017;117(11):1581-1582.
224. Eliceiri KW, Berthold MR, Goldberg IG, et al. Biological imaging software tools. *Nature methods*. 2012;9(7):697-710.
225. Niederlein A, Meyenhofer F, White D, Bickle M. Image analysis in high-content screening. *Comb Chem High Throughput Screen*. 2009;12(9):899-907.
226. Ridler TW, Calvard S. Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man, and Cybernetics*. 1978;8(8):630-632.

227. Huang LK, Wang MJJ. Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition*. 1995;28(1):41-51.
228. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber.* . 1979;9(1):62-66.
229. Zack GW, Rogers WE, Latt SA. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry* 1977;25(7):741-753.
230. Rosin PL. Unimodal thresholding. *Pattern Recognition*. 2001;34(11):2083-2096.
231. Riordan DP, Varma S, West RB, Brown PO. Automated Analysis and Classification of Histological Tissue Features by Multi-Dimensional Microscopic Molecular Profiling. *PLoS One*. 2015;10(7):e0128975.
232. Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy*. 2006;224(Pt 3):213-232.
233. Manders EMM, Verbeek FJ, Aten JA. Measurement of co-localization of objects in dual-colour confocal images. *Journal of microscopy*. 1993;169(3):375-382.
234. Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Human molecular genetics*. 2001;10(7):657-662.