

**DEVELOPMENT OF PANCREATIC CANCER BIOMARKERS FROM
GLYCAN VARIATION ON SERUM MUCINS REVEALED
BY GLYCOPROTEOMICS METHODS**

By

Tingting Yue

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Cell and Molecular Biology

2010

ABSTRACT

DEVELOPMENT OF PANCREATIC CANCER BIOMARKERS FROM GLYCAN VARIATION ON SERUM MUCINS REVEALED BY GLYCOPROTEOMICS METHODS

By

Tingting Yue

The development of serological biomarkers for pancreatic cancer could improve the clinical control of this disease by optimizing the process of diagnosis and treatment. The current best blood test uses the marker CA19-9, which detects the total level of the carbohydrate antigen sialyl Lewis A. The clinical usefulness of CA19-9 is limited due to its occasional elevation in non-cancer conditions as well as its low levels in some pancreatic cancer patients. Therefore, in this dissertation work, we aimed to improve the detection of pancreatic cancer over the standard marker CA19-9.

We used a novel strategy to develop biomarkers. The central hypothesis supporting this strategy is that by measuring the particular glycan on selected protein, we can improve the detection of pancreatic cancer compared to using the measurement of protein or glycan alone. The strategy enables us to examine the relationship between the changes in glycan and protein levels. It got implemented with the aid of a high-throughput sandwich assay which is based on antibody microarray technology and assisted by glycan binding proteins. The specificities of the glycan binding proteins were determined from a motif-based analysis.

In the first half of the study, we focused on three mucin proteins, MUC1, MUC5ac and MUC16 and profiled their glycosylation changes that were associated with pancreatic cancer. The elevations in glycosylation fell into 6 major groups which were determined by the specificities of the detection glycan binding proteins. Despite the overall elevation for all 6 groups of glycans, distinct patterns of alteration were found among the three mucins. Indeed, the

complementary elevation of glycan identified on the carrier mucin improved the detection of pancreatic cancer more than using the protein detection alone. This part of the study suggests that the three mucins are carriers of glycosylations associated with pancreatic cancer.

In the second half of the dissertation, we focused on the glycan CA19-9 and studied its prevalence on the three mucins and other proteins. Our aim was to develop a biomarker panel to distinguish pancreatic cancer from pancreatitis. A total of 420 individual samples including late-stage cancer, early-stage cancer, pancreatitis, and healthy control were obtained from 3 independent institutions for these studies. A complementarity strategy was developed for panel selection. We developed a biomarker panel enhancing the detection of pancreatic cancer than using the standard CA19-9 assay. We also identified a novel carrier protein for CA19-9, Apolipoprotein B-100, which did not contribute to the elevation of total CA19-9 in pancreatic cancer. ApoB represents the group of CA19-9 carriers irrelevant to pancreatic cancer development. Altogether, this part of study highlights the necessity of examining CA19-9 on carriers associated with pancreatic cancer.

In conclusion, this research demonstrates the effectiveness of a platform for glycan biomarker development. The platform features include high-throughput measurements of specific glycan changes on particular proteins and a complementarity strategy for selecting biomarker panels. The panel identified in this dissertation enhances the detection of pancreatic cancer and provides advanced patient stratification for future marker development.

ACKNOWLEDGEMENT

I would first like to give my sincerest regards to my dissertation advisor Dr. Brian B. Haab for his guidance, training and support for my graduate study. Next, I want to take the opportunity to thank my dissertation committee, Dr. J. Justin McCormick, Dr. John L. Wang, Dr. Timothy R. Zacharewski, and Dr. John J. LaPres from Michigan State University for their input and supervision of my dissertation work. I also want to send my sincere regards to the Cell and Molecular Biology Program, which brought me to MSU 5 years ago so I have the opportunity to meet and work with these excellent people. Third, I would like to send my appreciation to all the current and former lab members in the Lab of Cancer Immunodiagnostics, as well as other colleagues at Van Andel Research Institute. Forth, I want to thank the National Institute of Health (NIH), National Cancer Institute (NCI), and Van Andel Research Institute (VARI) for their sponsorship in this biomarker research.

Last but not least, I want to express my deepest thanks to my mother, father, and grandmother in Nanning, China, and to my fiancé in Cary, North Carolina. This dissertation is dedicated to my family.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
CHAPTER 1 GENERAL INTRODUCTION	1
CHAPTER 2 MICROARRAYS IN GLYCOPROTEOMICS RESEARCH (REVIEW).....	7
2.1 Introduction	8
2.2 Glycan microarrays	10
2.2.1 Technological overview	11
2.2.2 Determining specificities of glycan-binding proteins.....	12
2.2.3 Anti-glycan immune responses characterized using glycan arrays	13
2.2.4 Microbiology applications	15
2.2.5 Enzymatic studies	16
2.3 Lectin microarrays.....	16
2.3.1 Technological overview	18
2.3.2 Studies of protein glycosylation.....	19
2.3.3 Cell surface glycosylation.....	20
2.4 Antibody-lectin sandwich microarrays	21
2.5 Summary	24
Figures	27
CHAPTER 3 GLYCAN VARIATION ON MUCINS FROM PANCREATIC CANCER	32
3.1 Abstract	33
3.2 Introduction	34
3.3 Results	36
3.3.1 Profiling cancer-associated glycans on specific proteins.....	36
3.3.2 Prevalence of glycan changes relative to core protein changes.....	39
3.3.3 Discrimination of cancer from control using glycan detection.....	40
3.3.4 Structural insights from lectin-binding profiles	41
3.4 Discussion	42
3.5 Materials and Methods	48
Figures	52
CHAPTER 4 CA19-9 ON SPECIFIC PROTEINS	61
4.1 Abstract	62
4.2 Introduction	63
4.3 Results	65
4.3.1 Profiling cancer-specific glycans on specific proteins.....	65
4.3.2 Investigation of a panel to improve detection sensitivity over CA 19-9	66
4.3.3 Consistent improvement in sensitivity in independent sample sets.....	70

4.3.4 Size distribution of CA 19-9 carriers in the subgroups.....	71
4.4 Discussion	72
4.5 Materials and Methods	76
Tables	80
Figures	82
CHAPTER 5 NEW CA19-9 CARRIER IDENTIFIED IN PANCREATIC CANCER	93
5.1 Abstract	94
5.2.Introduction	94
5.3 Results	96
5.3.1 Variation in the carriers of the CA 19-9 antigen in patient subpopulations	96
5.3.2 Identification of protein carriers of the CA 19-9 antigen	98
5.3.3 Prevalence of the CA 19-9 antigen on ApoB in each disease state	99
5.4 Discussion	102
5.5 Materials and methods	105
Tables	110
Figures	112
CHAPTER 6 SYNTHESIS AND FUTURE WORK	119
6.1 Monitoring cancer-associated glycan alteration on selected proteins	120
6.2 Biological and/or statistical selection of biomarker panel	122
6.3 Future work and conclusion remark	123
BIBLIOGRAPHY	124

LIST OF TABLES

Table 1 Sets of plasma sets used in the study.	80
Table 2 Performance of individual markers for cancer from pancreatitis in each sample set.	80
Table 3 Comparison of the performance of total CA 19-9 and the marker panel in each set.	81
Table 4 Mass spectrometry identification of CA19-9 carriers.	110
Table 5 Mass spectrometry identification of apolipoprotein B-100.	111

LIST OF FIGURES

Figure 1 Microarray formats for glycoproteomics research.	27
Figure 2 Complementary detection of protein and glycan levels on using antibody arrays.....	28
Figure 3 Practical, high-throughput processing of antibody microarrays.....	30
Figure 4 Protein and glycan detection on antibody arrays.....	52
Figure 5 Cancer-associations of glycan levels.....	54
Figure 6 Glycan elevations relative to core protein elevations.....	56
Figure 7 Discriminating cancer from control using glycan or protein measurements.	58
Figure 8 Cancer-associations of major structural features.....	60
Figure 9 Detection of total CA19-9 and CA 19-9 on individual proteins using antibody arrays. 82	
Figure 10 Distribution of total CA19-9 levels in pancreatic cancer and pancreatitis patients.	83
Figure 11 Selecting complementary markers.	84
Figure 12 Combined results for all markers tested in UP set.	86
Figure 13 Raw images of arrays from subgroups defined by total CA19-9.	88
Figure 14 Panel performance in additional sample sets.....	90
Figure 15 CA19-9 western blot of individual samples.	92
Figure 16 Detection of CA19-9 on pre-identified protein MUC1, MUC5ac, MUC16.	112
Figure 17 Schematic of experiment design.....	114
Figure 18 Detection of CA19-9 on selected proteins.....	115
Figure 19 Detection of Apo B as CA19-9 carrier.	116
Figure 20 Apo B as CA19-9 carrier in healthy, pancreatitis and pancreatic cancer groups.	117

LIST OF ABBREVIATIONS

AUC, area-under-the-curve; CA, cancer antigen; CEA, Carcinoembryonic antigen; CEACAM, carcinoembryonic antigen-related cell adhesion molecule; Gal, galactose; GalNAc, N-acetyl galactosamine; GlcNAc, N-acetyl glucosamine; MUC, mucin; ROC, receiver-operator characteristic.

CHAPTER 1 GENERAL INTRODUCTION

Pancreatic cancer remains one of the most difficult forms of cancer to treat. In 2010, it is estimated that 43,140 new cases and 36,800 deaths in the United States will be caused by this cancer. Pancreatic cancer leaves only 3-6% 5-year survival rate to the patients, and this situation has not been changed since 1975 [1]. Only 10-15% of patients present with a potentially removable tumor, and many of them experience recurrence after surgery. For the majority of patients with locally advanced or metastasis pancreatic cancer, their median survival time is 3-10 months [2]. The exact mechanism of the risk factors for pancreatic cancer are still unclear. Factors, such as smoking, chronic and hereditary pancreatitis, late-onset diabetes mellitus and family history were used to indicate a relative higher risk of pancreatic cancer for an individual [3].

The vast majority of pancreatic cancer is the exocrine type which means the cancer resembles the exocrine pancreas where digestive pancreatic juices are produced. Within this type, ductal adenocarcinoma is the most common and accounts for over 90% of all pancreatic cancers. Around 70-80% of pancreatic cancer occurs at the head of the pancreas [2]. In addition to the invasive pancreatic cancer, three types of precursor lesions have been defined. Pancreatic intraepithelial neoplasia (PanIN) is the most common lesion and by far the best-characterized histological precursor of invasive pancreatic cancer [4].

In general, most cancer cells acquire six major types of capabilities during their development, albeit through various strategies involving common or separated sets of factors [5]. These capabilities include activation of growth, insensitivity to antigrowth signals, evasion of apoptosis, unlimited potential of replication, sustained angiogenesis, and tissue invasion as well as metastasis. In pancreatic cancer, the cancer cells acquire the self-sufficiency of growth signals by altering the expression of growth factors and their receptors. More than 95% of pancreatic

cancers present an abnormal activation of KRAS and an elevation of the gastrin and cholecystokinin receptor, and around 70% of pancreatic cancers demonstrate an increased expression of epidermal growth factor receptor (EGFR) and an abnormal activation of Src [3]. Around 48% pancreatic cancers also express focal adhesion kinase (FADK), which work closely with integrin pathway to generate additional growth signals [3]. All these alterations trigger a subset of cellular process such as transcription and translation to enhance proliferation of the cancer cells. To avoid being suppressed by the antigrowth signals, the TGF- β and SMAD4 pathway are altered in pancreatic cancer cells. Around 50% of pancreatic cancers are found with inactivation of SMAD4 to abolish the TGF- β mediated tumor-suppressive function[3]. For the third type of capability, evading the apoptosis, pancreatic cancer cells increase the expression of SHH and IHH in the hedgehog pathway, as well as Akt in the PI3K-Akt pathway. In order to escape the control of senescence and acquire the limitless replicative potential, 95% of pancreatic cancers overexpress the telomerase, which is essential to maintain the length of telomeres-the countdown clock for senescence. The next capability, sustained angiogenesis is acquired through the upregulation of vascular endothelial growth factor (VEGF), which was observed in more than 90% of pancreatic cancers. The final capability acquired during tumorigenesis is the tissue invasion and metastasis. The detailed mechanism for this complicated process has not been fully understood, but it is believed to be through the changes of the physical cell-to-cell coupling as well as the activation of extracellular protease.

The detection of pancreatic cancer is a challenge and so far disappointing. Surgery is the only curable treatment for pancreatic cancer, however, only a small portion of patients are detected at the stage when surgery is still an option. This situation is due to the vague symptoms presented by pancreatic cancer and the poor availability of detection tools to catch this disease.

In the patient in whom pancreatic cancer is suspected, an imaging tool using multiphase and multi-detector helical computed tomography (CT) with intravenous administration of contrast material is used to diagnose and stage the patient [6]. CT is particularly useful to determine the resectability of pancreatic cancer. It has an accuracy of around 80% [7]. One of the major limitations of CT and other imaging tool such as magnetic resonance imaging (MRI) is their inability to distinguish chronic pancreatitis from pancreatic cancer [6]. In such cases, additional diagnostic procedures such as endoscopic ultrasonography (EUS)-guided fine needle aspiration and endoscopic retrograde cholangiopancreatography (ERCP) are used to obtain tissue from pancreas [6]. However, the invasive diagnostic procedures can result in severe complications such as pancreatitis and infection, occurred in roughly 5% and 15% of post-ERCP patients, respectively [8].

CA19-9 is the best serum marker for pancreatic cancer by far. It detects a carbohydrate antigen sialyl Lewis A [9], which is found on multiple proteins including some mucins.

Unfortunately, standard CA19-9 assay is of limited usefulness for therapeutic monitoring and assisting detection of post-surgery recurrence of disease [10, 11]. The limited use of CA19-9 is due to its elevation in non-pancreatic disease as well as the natural existence of a CA19-9-negative subgroup of pancreatic cancer patients [12]. Biomarker development for pancreatic cancer has been intensively focused in the past decades. Over 2500 genes were identified as producers of biomarker for this cancer, at either the DNA or the protein level [13]. Much less research has been done to explore the potential of post-translational modifications such as glycosylation as biomarkers for pancreatic cancer.

Glycans can present either in a free form, or associated with a particular scaffold to generate glycoproteins or glycolipids. The biosynthesis of a glycan, unlike other biopolymers, is

not template-driven. In mammals, nine monosaccharide units serve as building blocks for all glycan synthesis. The biosynthesis process is accomplished by the combined function of two major types of enzymes, glycosyltransferases and glycosidase, for addition and removal of glycan unit, respectively. This combined action leads to the production of various types of glycan structures. On glycoproteins, the glycan structures can be grouped in one of the 3 major classes. N-link glycans are attached to the asparagine residue of a carrier protein, while O-link glycans are attached to the serine and threonine residues. O-link glycans are sometimes referred to mucin-type O-link glycans because of the extensive O-glycosylations found on the mucin family of glycoproteins. The third type, glycosaminoglycans (GAGs) are linearly attached to the serine residues of proteoglycan molecules. There are an enormous number of possible combinations of the glycan structures, synthesized by the cells to initiate or in response to a particular process.

With the advance in knowledge of glycobiology, the association between cancer and altered glycosylation on proteins is increasingly clear [14]. Glycosylation changes on both the tumor surface and host elements, to mediate the key pathological events required for tumor progression in various steps, such as invasion, proliferation and angiogenesis [14]. These alterations include both the aberrant amount or type of glycan chains as well as the altered expression of the glycoproteins such as mucins.

Compared to normal cells, the complexity of glycosylation is clearly increased in tumor cells, through elevated production of glycans with both increased and decreased size. One typical increased-size change in cancer is the increased N-glycan branching resulted from the overexpression of the N-acetylglucosamine transferase V (GlcNAc T5). On the other hand, tumor cells also produced decreased-size change, the truncated O-glycans. Examples include sialyl Tn antigen or sialyl Tn antigen, both containing less than three monosaccharide units. The

production of truncated glycan chains may involve the competition among glycotransferases responsible for O-glycan production. Furthermore, addition of sialic acids to the T and Tn antigen prevents the further elongation of the truncated structures. In pancreatic cancer, the most frequent glycan change identified so far is the overexpression of CA19-9. CA19-9 is an important component of the natural ligands for the endogenous selectins. Their interactions play a part in the metastatic cascade of the pancreatic cancer cells, probably through the interaction with leukocytes to facilitate cancer cell transportation to ecopic sites and evasion of the immune system [15].

With the help of glycan-binding-proteins such as carbohydrate-antibodies and lectins, glycan alterations can be well detected. Alteration of glycan chains on protein provides an additional dimension of molecular modification to meet the cellular needs during cancer development. Exploration on these glycosylation changes leaves us a great opportunity to further understand, and more importantly, better control the disease.

This dissertation research utilized microarray-based glycoproteomics tools to develop glycan biomarkers for pancreatic cancer detection. In the following chapters, I will present 1) a review of microarray application in glycoproteomics research, 2) the prevalence of glycosylation alteration in pancreatic cancer on targeted proteins revealed by profiling lectin bindings, 4) the development of biomarker panel enhancing the detection of pancreatic cancer, and 5) the identification of a novel CA19-9 carrier and the lesson from it. The ultimate goal of this study is to deliver the benefits of bench science to the patients.

CHAPTER 2 MICROARRAYS IN GLYCOPROTEOMICS RESEARCH (REVIEW)

T. Yue and B.B.Haab (2009). Microarrays in glycoproteomics research. *Clin Lab Med.* 29, 15-29.

2.1 Introduction

The importance of carbohydrate post-translational modifications on proteins has become increasingly clear over many decades of advances in our understanding of glycobiology. It is now widely appreciated that the carbohydrate side chains play critical roles in the structure and function of many, if not most, cell surface and secreted proteins. These roles include the guiding of proper protein folding, the maintenance of protein conformation, mediating receptor-ligand and protein-protein interactions, providing biophysical polarity or hydrophilicity, and guiding immune recognition [15]. Abnormal glycosylation also is associated with a variety of inherited and sporadic diseases [14, 16], testifying to the necessity of proper glycosylation for the maintenance of health. Furthermore, glycosylation has been found to be an integral part of the proper functioning of every organism in nature. Given this broad importance of glycobiology, the field has increasing importance in applied research in biotechnology and biomedicine. For example, treatment strategies based on interfering with glycan-mediated process or targeting cancer glycans are under development [17], and blood-based diagnostic tests using glycan detection may be possible [14].

Protein glycosylation refers to the chains of monosaccharide building blocks that are covalently linked to particular amino acid residues (usually serine, threonine, or asparagine residues). The monosaccharides are often five- or six-carbon cyclic structures with various modifications and isometries. Glycosidic bonds join the monosaccharides at any of the carbons with either an ‘alpha’ or a ‘beta’ linkage—referring to the stereoisometry of the linkage at a chiral carbon—and in a linear or branched fashion. This variety in the components and the linkages results in a huge diversity of structures that can be formed. In reality, although great

diversity indeed is observed in each organism, particular motifs and structural themes are prevalent, bringing some order to the complexity.

The study of carbohydrate biology has been viewed as daunting by many researchers due to the heavy chemistry emphasis and perceived difficulty in the traditional analytical methods. Indeed, historic glycobiology has focused primarily on carbohydrate chemistry and involved techniques not typically included in the education of biologists. However, the development of modern biological methods provided new ways of studying glycobiology, resulting in huge advances in our understanding of the genetic and biochemical basis of glycan synthesis and the molecular and cellular biology of glycans. The new tools for studying glycobiology make the field more accessible and useful to a broader base of researchers—tools such as glycan-binding antibodies and proteins, genetically-modified cells and organisms, gene expression profiling of glycan-related genes, and advanced chromatographic and mass-spectrometry analysis of glycan structures.

Another set of tools that is advancing glycobiology is built on the microarray platform. Microarray methods analyzing RNA and DNA transformed gene expression and genetic research following their introduction in the early 1990s [18]. The usefulness of the microarray platform lies in its multiplexing capability—enabling the acquisition of many data points in parallel—and its miniaturization—resulting in very small consumption of reagents and samples. These benefits were recognized by researchers studying other molecule types, including proteins, antibodies, lipids, and glycans, but microarrays for such studies developed more slowly due to increased technical difficulty. Currently, microarrays for studying all types of molecules are becoming established and broadly applied.

This article deals with the application of microarray formats to the study of glycoproteins and glycans. We survey the technology and applications of three different types of microarrays developed for glycoproteomics research: glycan microarrays, lectin microarrays, and antibody-lectin sandwich microarrays (Figure 1). Each format is used in distinct and complementary types of experiments and has facilitated advances in glycobiology research. The review focuses on glycans and proteins and therefore will not cover DNA arrays, such as the Glyco-gene DNA chip (provided by the Consortium for Functional Glycomics) for profiling the expression of genes involved in the glycosylation machinery.

2.2 Glycan microarrays

A major goal in glycobiology research is to probe and characterize interactions between glycans and various types of glycan-binding proteins. Conventional methods for such studies are not suitable for the profiling of many different glycans (“glycomics” studies). For example, the glycan-binding specificities of lectins have been effectively probed by determining the elution profiles of various glycans in affinity chromatography [19], but these experiments require significant amount of glycan material per test with each glycan tested sequentially. Likewise, enzymatic studies on glycans can only be performed on a single glycan per assay and require significant material.

The glycan microarray addresses these limitations by enabling binding analyses to many different carbohydrate structures in small sample volumes and with minimal consumption of the reagents. The low consumption of the carbohydrate structures is particularly important because of the difficulty and time required to synthesize or isolate those structures. Below we present an

overview of the types of carbohydrate microarrays that have been produced, follow by a review of the applications of carbohydrate microarrays.

2.2.1 Technological overview

Several different demonstrations of carbohydrate microarrays appeared in 2002 with a variety of fabrication techniques [20, 21]. Houseman et al. used covalent attachment of carbohydrate-cyclopentadiene conjugates to self-assembled monolayers on gold surfaces to create monosaccharide arrays, and these chips were used to test interactions with selected lectins and enzymes [22]. Mono- and di-saccharide chips were created by Park and Shin [23] using the covalent attachment of maleimide-linked carbohydrates to thiol-derivatized slides. Fazio et al. attached oligosaccharides to hydrocarbon chains, which were then attached in a non-covalent fashion to the bottoms of wells of polystyrene microtiter plates [24]. These oligosaccharides were probed with lectins and also were modified directly in the plate using glycosyltransferases. Wang et al. used non-covalent attachment to create glycan microarrays, by spotting carbohydrate-containing macromolecules onto nitrocellulose-coated glass slides [25]. A wide variety of microbial antigens were probed with anti-carbohydrate antibodies. Fukui and colleagues also used a nitrocellulose surface, onto which they spotted lipid-linked oligosaccharide probes generated from glycoproteins and polysaccharides [26]. Non-covalent attachment also was used by Willats and coworkers, in which complex polysaccharides, proteoglycans, and neo-glycoproteins were spotted onto oxidized polystyrene, followed by probing with anti-glycan antibodies [27]. Several of the applications above show the value of oligosaccharides derived from biological sources, which represent complex, biologically-relevant structures that are not able to be synthesized. Xia et al. also demonstrated a practical method for

the generation of such probes [28]. The above demonstrations primarily used fluorescence scanning of dye-labeled probes for detection.

A significant advance in the utility and availability of glycan microarray technology came through the development of glycan microarrays by the Consortium for Function Glycomics (CFG). The CFG is dedicated to providing resources, technology and collaborative opportunities for investigators focused on carbohydrate-related research. Researchers from the CFG synthesized over 200 biologically-relevant glycans attached to amine-conjugated spacers, and spotted them onto NHS-activated glass slides to form covalent linkages [29]. The arrays were initially used to characterize specificities of plant lectins, human lectins, glycan-binding antibodies, and bacterial and viral proteins [29]. Since then, the CFG has profiled the specificities of hundreds of glycan-binding proteins for researchers participating in the consortium, and the data are made available through the CFG website. Some of the applications of these and other glycan microarray platforms are described below.

2.2.2 Determining specificities of glycan-binding proteins

The most widely-used application of glycan microarrays is to characterize the specificities of glycan-binding proteins and antibodies. An important class of molecules for which glycan-binding specificity has been studied is lectins, which are non-enzymatic, glycan-binding proteins found in all types of organisms. Lectins play roles in diverse processes such as cell migration, immune recognition, and angiogenesis, among others [30], and information about lectin binding specificities gives important clues about function. In addition, the characterization of lectin specificity is helpful for their use as analytical reagents.

Manimala et al. [31] developed a glycan microarray containing 54 conjugated glycans, ranging from basic monosaccharides to more complicated cancer associated glycan motifs, to

obtain the binding profiles of 24 lectins. Glycan microarray slides were incubated with biotinylated lectins in serial dilutions and probed with streptavidin-horseradish peroxidase. The authors identified specificities that were not previously observed due to the lack of a practical approaches for screening that many interactions. For example, a similar study using the conventional ELISA method would require the use of over 100-fold more sample.

Glycan microarrays also have been useful to characterize the specificities of monoclonal antibodies raised against carbohydrates. Microarrays containing 80 different carbohydrates were used by Manimala and coworkers to study the specificities of 27 different anti-glycan antibodies, which showed that most antibodies bound other glycans in addition to their nominal targets [32]. This finding may reflect the difficulty in generating truly specific anti-glycan antibodies. The same group also showed that antibodies raised against the Tn antigen (GalNAc α 1-O-Ser/Thr), an important epitope in cancer and microbiology, also reacted with the blood group A structure (a related glycan), which has implications for diagnostics using anti-glycan antibodies [33]. In related applications, the glycan-binding specificities of antibodies raised against the cell wall polymers from the plant *Arabidopsis thaliana* were characterized using arrays containing 50 cell wall glycans [34], and specialized microarrays were designed to characterize the specificities of monoclonal antibodies targeting the Globo H hexasaccharide found on cell surfaces from breast, prostate, and ovarian cancers [35].

2.2.3 Anti-glycan immune responses characterized using glycan arrays

Certain glycan structures can elicit an immune response if the structures are not normally presented to the host immune system. Indeed, the blood group glycan structures of the ABO system elicit antibodies against structures not found in the host, which can lead to the agglutination of red blood cells from unmatched blood. Anti-glycan antibodies also can be

generated in immune responses against pathogens and cancer. The glycan array provides a powerful tool for probing for anti-glycan antibodies and for determining their specificities. In the work of Lawrie et al. [36], a glycan microarray containing 37 covalently bound glycans was used to identify carbohydrates that trigger humoral immune responses in classical Hodgkin's lymphoma (cHL) patients. Total IgG and IgM from groups of patient and age/sex matched control individuals were purified, pooled within the same diagnostic states, and applied to individual arrays. Biotin-conjugated anti-human IgG or IgM antibodies, followed by dye-labeled streptavidin and fluorescence scanning, were used to detect anti-glycan antibodies. Antibodies against five carbohydrates, including the T and Tn antigens (Gal β 1,3GalNAc α - and GalNAc α -, respectively) were identified from cHL patients and became targets for further investigation.

In another cancer-related study, glycan microarrays were used by Wang et al. to show that breast cancer patients have an increased level of antibodies targeting a glycan structure called Globo H, which is found on the surfaces of breast cancer cells [37]. Globo H and its truncated analogs were robotically spotted at various concentration onto NHS-coated glass slides for covalent attachment, and bound antibodies were detected by Cy3-labeled anti-IgG secondary antibodies. The detection of anti-Globo H antibodies may be useful for breast cancer diagnostics.

Antibodies against bacterial pathogens also were detected using carbohydrate arrays. Arrays containing glycans from anthrax toxin (*B. anthracis*) were generated to characterize antibodies from rabbits that had been infected with anthrax [38]. Antibodies were found that reacted with a prominent tetrasaccharide specifically found on the spore surface, which confirms this structure as an important immunological target. A carbohydrate microarray containing oligosaccharides specific for different versions of *Salmonella enterica* was used to identify glycan-reactive antibodies and to confirm the presence of strain subgroups in patients suffering

from salmonellosis [39]. This method might be useful for rapid diagnosis or for epidemiological or vaccine studies.

2.2.4 Microbiology applications

Glycans play important roles in the recognition and attachment to host sites by microbial pathogens. In most cases the nature of the glycan attachment sites is not known. Glycan microarrays have been useful for studying that question. Stevens et al. [40] used glycan microarrays to investigate the host specificity of influenza viruses. Glycans with different carbohydrate components or glycosidic linkages printed in the microarray were incubated with recombinant hemagglutinins from membranes of various influenza virus strains or hybridized directly with whole virus. Different virus strains were found to have distinct preferences of binding to particular glycosidic linkage types. Profiling of strain-specific glycoprotein binding specificity from various virus strains using glycan microarray could provide crucial information in understanding virus adaptation and species barriers, which may be useful for preventing human infection.

Disney et al. [41] used glycan microarrays containing five different monosaccharides to detect pathogens and test for their antibiotic susceptibility. Microarray slides were hybridized with *E. coli* cells that had been labeled with a nucleic acid staining dye. After washing away unbound bacteria, slides were scanned to detect fluorescence, indicating the glycan-bacteria binding. Mutant strains with altered carbohydrate binding patterns could be detected from this assay. The authors proposed that a carbohydrate-binding “fingerprint” identified using this glycan microarray can be used to determine the types of bacteria present within a complex mixture. Since this is a nondestructive method, bacteria captured on the arrays can further be

harvested and tested for antibacterial susceptibility, which is not possible using traditional destructive methods, such as those requiring PCR.

2.2.5 Enzymatic studies

Another important application of glycan microarrays is to evaluate the specificities and activities of sugar-processing enzymes, such as transferases used for the addition of carbohydrate units [42-44]. Increased knowledge about glycosyltransferases and glycosidases would lead to a better understanding of how certain glycan chains are produced and how to use these enzymes for the synthesis of carbohydrates. Several groups have demonstrated the modification of sugars immobilized in arrays. In the work of Park et al. [43], the treatment of glycan arrays with UDP-Gal and β -1,4-galactosyltransferase (GalT) resulted in the conversion of N-acetylglucosamine (GlcNAc) to lactosamine only in the absence of fucose on the GlcNAc, thus providing valuable information about the enzyme specificity. In another experiment from the same group, the authors successfully synthesized Sialyl Lewis X ($\text{NeuNAc}\alpha 2,3\text{Gal}\beta 1,4(\text{Fuc}\alpha 1,3)\text{GlcNAc}$) from arrayed GlcNAc by the addition of a series of glycotransferase and sugar units [42]. Their work showed the efficiency and potential of using glycan microarray to characterize carbohydrate-processing enzymes as well as the possibility of enzymatic transformation from simple glycans to complex carbohydrates directly on the array. In more recent work, the acceptor specificities of multiple sialyltransferases were compared, showing differences in substrates between human and rat sialyltransferases [44].

2.3 Lectin microarrays

Lectin microarrays also take advantage of the low-volume and multiplexing capabilities of microarrays, but provide complementary information to glycan microarrays. Lectins were first

recognized by their ability to agglutinating red blood cells [45], and later the term “lectin” was adopted when it was realized that there existed a class of carbohydrate-binding proteins [45]. Although lectins were originally isolated from plants, they were later found ubiquitously in nature [46]. Lectins originally were classified according to their glycan-binding specificities, but they are now more consistently grouped according to sequence and structural motifs [47].

Plant lectins have become extremely valuable analytical tools for the detection of particular carbohydrate structures. As affinity reagents, lectins can detect glycans with high reproducibility and in a variety of formats, and thus provide a good alternative to other glycan detection methods involving enzymatic digestion, chromatography or mass spectrometry, which have low reproducibility and low throughput. The specificity of detecting particular glycans depends on the lectin used; some are highly specific, and others have broad specificities. Lectin detection can have good sensitivity due to multi-valent binding, resulting from either a multi-subunit or multiple carbohydrate-binding sites within a single polypeptide [48].

Lectins have long been used for detection and purification of glycans in various research fields [49]. Lectins have been used extensively in immunohistochemistry, for example in studies to examine the tissue distribution in pancreatic tumors of certain blood-group carbohydrates [50, 51]. Lectins also have been used in immuno-affinity electrophoresis and blotting methods, for example to identify cancer-associated glycan variants on the serum proteins α -fetoprotein [52], haptoglobin [53, 54], α -1-acid glycoprotein [55], and α -1-antitrypsin [56]. Although lectins have been used as sugar detection reagents for decades [30], the use of multiple lectins for the analysis of biological samples was not common due to the amount of material and time required for each assay.

2.3.1 Technological overview

The lectin microarray made it practically feasible to obtain glycan measurements on a given sample from multiple, different lectins. By incubating samples on an array of lectins and determining the amount of binding to each lectin, a broad profile of the glycans present in the sample can be rapidly obtained with minimal sample consumption. This approach has many advantages over standard methods of glycan analysis, such as reduced cost, time, and sample consumption, with increased reproducibility. An additional advantage is that lectins can provide information about linkages between monosaccharides (for example whether the alpha or beta configuration), which is not discernable using mass-spectrometry analysis.

Some challenges accompany the use of lectin microarrays. One is the question of specificity. Lectins cannot give exact structural information, but rather mainly provide information about terminal structures. Moreover, the specificities of some lectins are incompletely understood and may involve more than one particular structure. Some have suggested that through the use of many lectins, the analysis of an overall lectin-binding profile can overcome some of the ambiguities associated with lectin binding [57], although the effectiveness of such a strategy has not been demonstrated. Another challenge can be detection sensitivity, since lectin-glycan interactions are weak compared with DNA-RNA or antibody-antigen hybrids. Standard washing protocols developed for protein or DNA microarrays may result in the significant loss of binding. An evanescent-field fluorescence strategy is a promising solution [58], since it allows the sensitive, real-time observation of monovalent lectin-oligosaccharide interactions under equilibrium conditions without the requirement of washing. The optimization of this method enabled the detection of glycoproteins down to a 100 pM concentration [59].

The initial demonstrations of lectin microarrays [60, 61] used standard arraying methods that had been developed for DNA and protein microarrays. Lectins were printed on aldehyde- or epoxide-derivatized glass slides to achieve covalent immobilization, or were linked via biotin-streptavidin bridges on photoactivatable dextran-coated slides. Another lectin array used the non-covalent suspension of lectins in an aqueous hydrogel matrix [62], which may better preserve the conformation and activity of lectins relative to binding on planar surfaces. The arrayed lectins were hybridized with samples (e.g. glycoproteins) that were directly labeled with a fluorescent tag or that were recognized by a labeled detection reagent.

2.3.2 Studies of protein glycosylation

The major application of lectin microarrays has been to rapidly investigate the glycosylation of purified glycoproteins which were incubated on the arrays. For example, Kuno and coworkers used arrays containing 39 lectins to detect glycosylation differences between various glycoproteins and changes in glycosylation induced by treatment with glycosidases [58]. The incubation of purified proteins, as opposed to mixtures of proteins, is important to simplify the interpretation of the data, so that one may know the identity of the protein binding each lectin. However, others have demonstrated the incubation of complex mixtures of proteins onto lectin arrays, thus achieving a summary view of a cell “glycome.” Pilobello and coworkers used a ratiometric approach to examining changes in bacterial cell-surface glycomes [63]. Isolated membrane proteins from two bacterial cultures were differentially labeled with Cy3 and Cy5 fluorescent dyes and co-incubated on arrays containing up to 58 different lectins. The Cy3/Cy5 ratio at each spot provided a sensitive indicator of differences between the cultures and allowed for normalization between arrays. This analysis enabled the observation of glycosylation changes occurring in response to cell differentiation. The evanescent-field fluorescence method

mentioned above also was applicable to the study of crude glycoproteins extracted from mammalian cells [64]. While the approach of incubating multiple proteins on lectin arrays offers a summary view of the glycan structures on a cell, it has the disadvantage of integrating information from all proteins, so that glycan changes that occur only on a subset of proteins may be lost in a background of non-changing proteins.

2.3.3 Cell surface glycosylation

A potentially simpler and more direct view of cell-surface glycosylation has been achieved by incubating live cells on the surfaces of lectin microarrays. The use of whole cells as opposed to cell extracts has an advantage of preserving higher-order structures, which may be biologically significant and important for lectin binding. Early work by Zheng et al. [65] used covalent immobilization of lectins on self-assembled monolayers that were functionalized with NHS. Cultured cells were incubated on the spotted lectins, and the binding of the cells to the lectins was visualized with an inverted microscope. The gold base substrate was thin enough to allow the imaging. The authors showed differences in the glycosylation of the two cell types. In later work by the same group [66], the authors used this technology to explore glycan differences between normal and breast cancer cell lines. Significant variation in glycosylation was identified which correlated with metastatic potential as well as metastatic location preference.

Lectin microarrays also were used to examine dynamic changes to *E. coli* bacterial glycosylation [67]. The bacteria were labeled with a dye that binds to DNA to allow detection by fluorescence after incubation on the arrays. The authors could distinguish *E. coli* strains based on glycosylation and could observe growth-dependent variation in glycosylation on particular strains. Lectin arrays employing evanescent-field fluorescence, as described earlier [58], were used to examine dynamic changes to the cell surfaces glycomes of mammalian cells that had

been fluorescently labeled with a DNA-binding dye [68]. Alterations in lectin-binding patterns were seen in glycosylation-defective mutants of CHO cells and in splenocytes from mice with a genetic knockout of a glycosyltransferase gene. Changes in cell surface glycosylation associated with erythroblast differentiation also were observed. Another study using arrays of 94 lectins and a similar detection method examined the lectin-binding signatures of 24 different human cell lines and predicted functional phenotypes based on lectin-binding profiles [69].

2.4 Antibody-lectin sandwich microarrays

Another array-based glycoproteomics method is the antibody-lectin sandwich microarray. The value of antibody-lectin sandwich microarrays for glycoproteomics studies is that they provide precise measurements of glycan levels on specific proteins captured directly from biological samples. This capability enables detailed views of how glycans on particular proteins change in association with disease states or sample conditions. Previous methods did not practically allow that type of investigation. Studies employing enzymatic, chromatographic, and mass spectrometry methods have been very effective for providing detailed information about glycan structures in individual samples, but due to high sample consumption, low throughput, or low reproducibility, such studies did not reveal how frequently particular glycans on particular proteins appear, how closely they are associated with particular disease states, or the distribution of protein carriers on which they appear. Affinity-based methods, using reagents such as lectins or glycan-binding antibodies to detect glycans, can provide that information, because one may reproducibly measure the glycan levels over multiple samples. While affinity-based glycosylation studies do not provide the structural detail provided by mass spectrometry and

enzymatic methods, they can provide information about the biological variation of a particular motif.

The method starts with an antibody microarray—essentially identical to those developed for multiplexed protein analyses [70] (Figure 1). The antibodies on the array can be chosen to target various glycoproteins. A complex biological sample is incubated on the array, resulting in the capture of glycoproteins by the antibodies. The next step is to probe the glycans on the captured proteins using labeled lectins. The amount of lectin binding at each capture antibody indicates the amount of a particular glycan on the proteins captured by each antibody. A variety of lectins could be used on a given sample in order to probe several different types of glycans. Glycan-binding antibodies also could be used as detection reagents, such as those raised against the Thomsen-Friedenreich antigens [71] or the Lewis blood-group structures [72]. Antibody-lectin sandwich arrays are similar to previous approaches using lectins in the capture or detection of proteins in microtiter plates [73], but they harness the power of microarrays to provide high information content in low sample volumes.

In order to properly interpret the amount of glycan on a protein, one must also know the underlying protein concentration. That complementary information may be conveniently obtained using antibody microarrays in a standard sandwich assay format to detect core protein levels (Figure 2). Therefore a sample may be incubated multiple times on replicate microarrays, each time probed with a different lectin, to characterize glycan levels (Figure 2b), or with antibodies, to characterize protein levels (Figure 2a). A previous study [74] showed the value of using both formats to detect glycosylation differences between samples (Figure 2c).

The ability to probe each sample multiple times, and to probe many samples (as would be required for clinical studies), requires the ability to run many samples efficiently and to consume

small sample volumes in each assay. A practical method for the high-throughput processing of low-volume microarrays was demonstrated earlier [75]. Multiple, replicate microarrays are printed onto a microscope slide, and the arrays are separated from one another by hydrophobic, wax borders that are precisely imprinted onto the slide. The borders are imprinted using a device (The Gel Company, San Francisco, CA) that elevates a stamp out of a wax bath, which sits atop a hotplate to melt the wax, to contact a microscope slide suspended above the wax bath (Figure 3a). The wax borders prevent liquid from spilling from one array to another, and they remain on the slide throughout the processing steps and the fluorescence scanning (Figure 3b). Any size or pattern of arrays could be accommodated by using the appropriate stamp (Figure 3c). A high-throughput strategy using this format is to incubate sets of samples in a randomized order on a microscope slide and then probe the captured proteins on the slide with a lectin (Figure 3d). Such a strategy has been used for high-throughput antibody array processing in multiple studies [74, 76, 77].

Ongoing studies are applying this technology in various ways. One application is the profiling of glycan changes on particular proteins in various disease states, to identify those most associated with disease and to develop new biomarkers. Unpublished data in our lab shows that certain glycans on particular serum proteins are altered in pancreatic cancer more frequently than the underlying core protein. Because of that relationship, the measurement of the glycan on the protein performs better as a biomarker than the measurement of the protein. The same study also provides information on the prevalence of glycan alterations in various patient populations, which was not known before because of the aforementioned limitations of conventional glycobiology methods. Another application of this technology is the study of glycan changes induced by various perturbations to cultured cells. For example, we have examined the effects of

pro-inflammatory stimuli on the glycan structures of mucins secreted by cancer cells. We found that the induced glycan structures are similar to those observed in clinical samples from pancreatic cancer patients, suggesting that cancer-associated glycans can arise in response to a pro-inflammatory tumor microenvironment. These types of studies show how the capabilities of antibody-lectin sandwich arrays, such as the ability to measure specific glycans on multiple, specific proteins captured from biological samples, and the ability to precisely measure changes between samples in those glycan levels, enable access to information that would be difficult to acquire using conventional methods.

2.5 Summary

The above overview provides insights into how microarray platforms are stimulating advances in glycoproteomics research. Each of the platforms discussed here can be used in multiple types of experiments to produce distinct types of information. The many different areas of glycobiology in which glycan, lectin, and antibody microarrays have been applied testify to the versatility of the platforms. Glycan arrays are valuable for studying protein and cell binding to glycan structures, and they have been used to study glycan-binding specificities, disease-associated anti-glycan antibodies, microbe-carbohydrate interactions, and the activities of glycan-processing enzymes. Lectin arrays provide a convenient and rapid tool for assessing the overall glycan content of a sample. These tools have facilitated studies on the glycosylation of purified proteins and complex mixtures of proteins as well the total carbohydrate content of cell surfaces. Finally, antibody-lectin sandwich microarrays are an additional tool for glycoproteomics, and provide information about the carbohydrate on multiple, distinct proteins captured directly from biological samples. This tool enables views of glycan

variation on specific proteins in patient populations or in response to changing conditions. All of these tools can provide experimental information that was not obtainable using conventional technologies. Therefore, the increased usage of these approaches is anticipated to drive further major advances in glycoproteomics.

APPENDICES

FIGURES

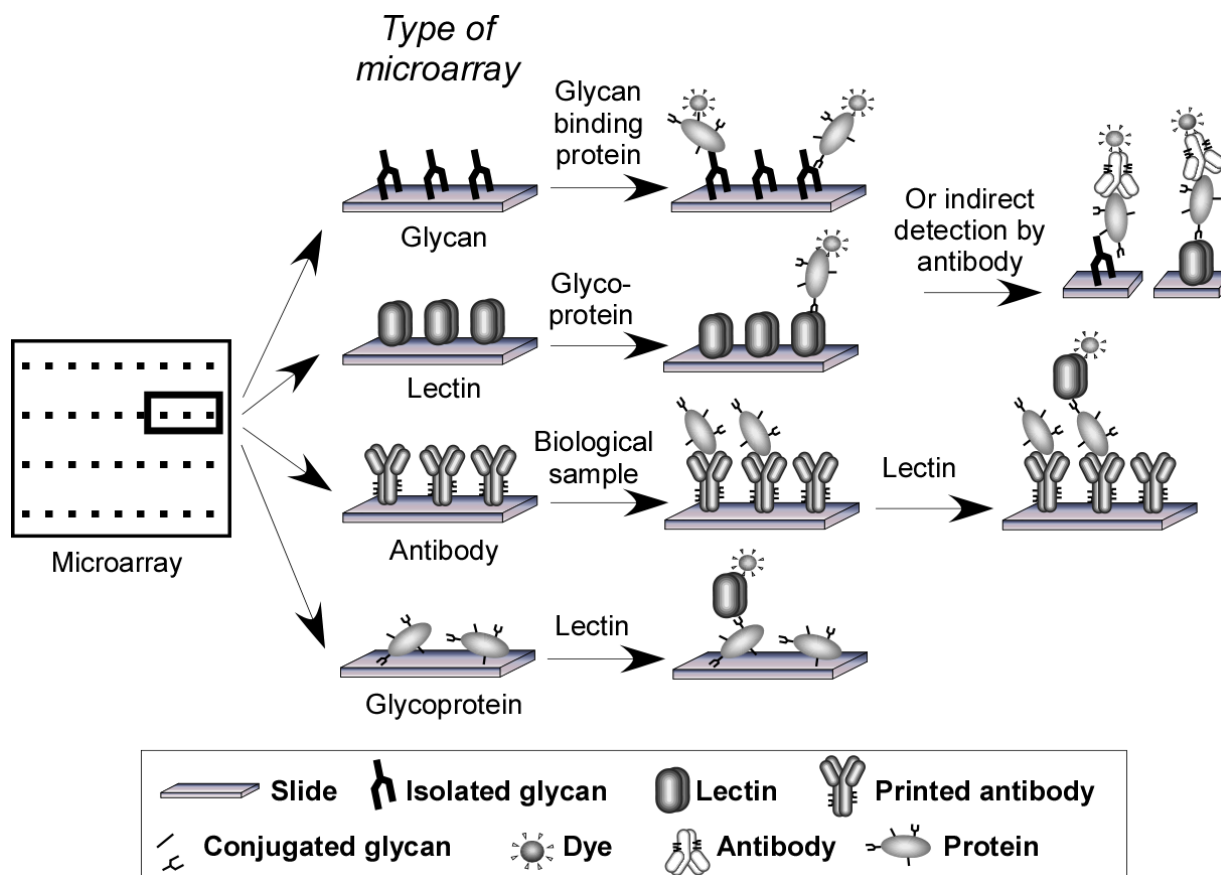


Figure 1 Microarray formats for glycoproteomics research.

The types of microarrays depicted are glycan arrays, lectin arrays, antibody-lectin sandwich arrays, and glycoprotein arrays. A detection strategy using a fluorescent dye is depicted, although other detection methods could be used, such as surface-plasmon resonance or chemiluminescence. Glycoprotein arrays, involving the isolation of glycoproteins from biological samples using chromatographic methods, followed by the probing by lectins of the glycans microarrayed proteins [78, 79], are not discussed in the text but are depicted here for completeness.

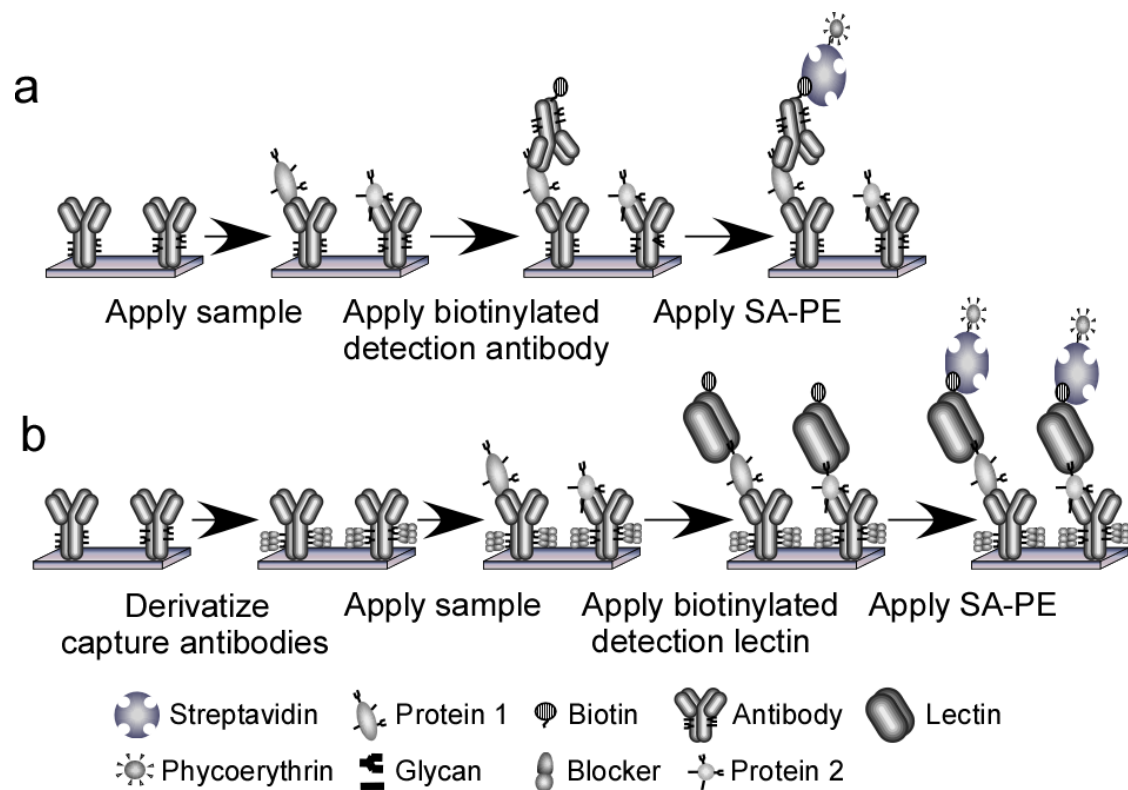


Figure 2 Complementary detection of protein and glycan levels on using antibody arrays.

a) Array-based sandwich assays for protein detection. Multiple antibodies are immobilized on a planar support, and the captured proteins are probed using biotinylated detection antibodies, followed by fluorescence detection using phycoerythrin-labeled streptavidin. b) Glycan detection on antibody arrays. This format is similar to above, but the detection reagents target the glycans on the captured proteins rather than the core proteins. The glycans on the immobilized antibodies are chemically derivatized to prevent lectin binding to those glycans. c) Detection of differential glycosylation. A healthy patient serum sample and a cancer patient serum sample were incubated on each pair of arrays (same two samples in each pair). The boxes indicate the capture antibodies targeting MUC1 and CEA. The arrays were detected using either a mixture of two antibodies to detect the MUC1 and CEA core proteins (left

array in each pair) or an antibody detecting the CA 19-9 carbohydrate epitope (right array in each pair). The data show equivalent core protein levels between the healthy and cancer serum, but high levels of the CA 19-9 glycan in the cancer serum. The other spots showing signal are control proteins or other proteins containing the CA 19-9 epitope.

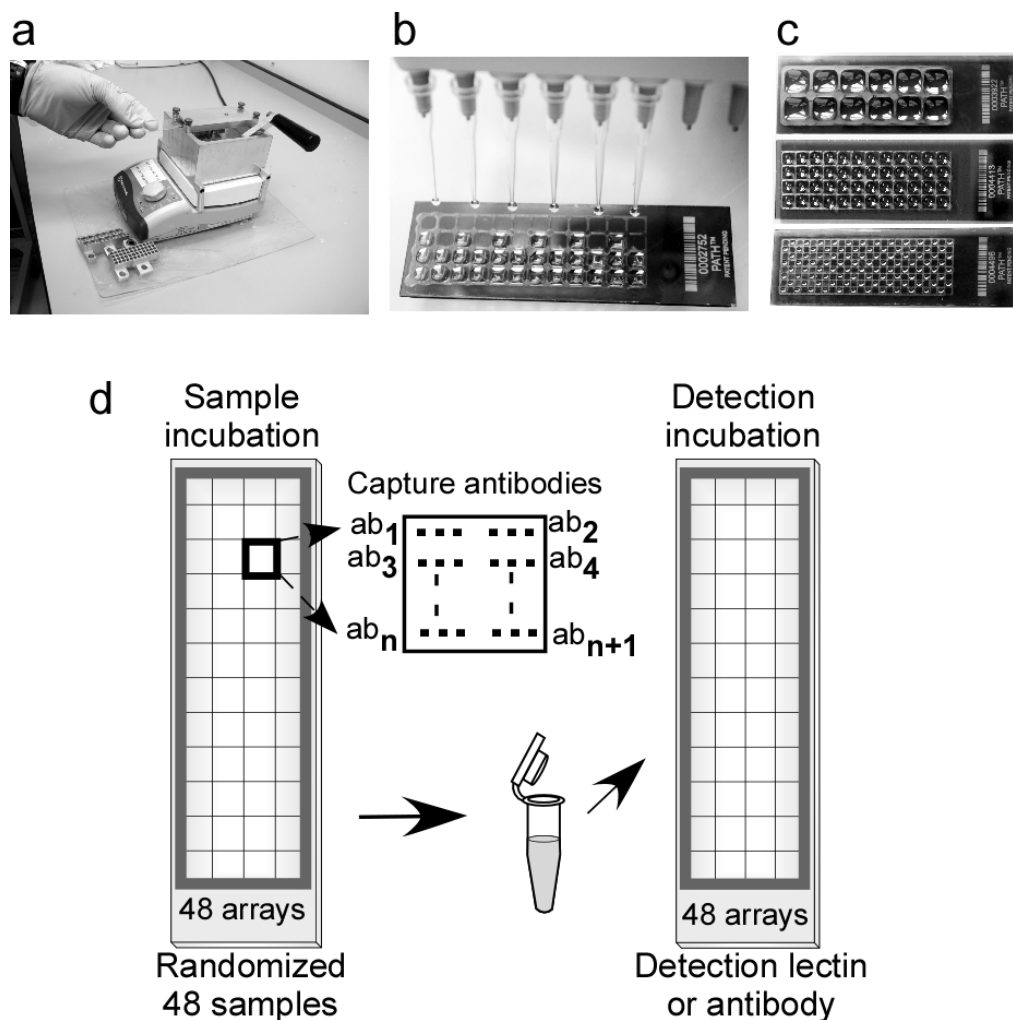


Figure 3 Practical, high-throughput processing of antibody microarrays.

a) Imprinting hydrophobic boundaries. Wax is melted by the hotplate under the bath, and a slide is inserted upside-down into the holder. Bringing the lever forward raises a stamp out of the wax bath to touch the slide, imprinting the design onto the slide to form borders around multiple arrays. Two stamps are shown in front of the machine. b) Loading samples onto a slide containing 48 arrays. The arrays are spaced by 4.5 mm, which is compatible with the 9 mm spacing of standard multi-channel pipettes. c) Samples loaded onto slides containing 12 (top), 48 (middle), and 192 (bottom) arrays (96 samples loaded). d) Strategy for profiling glycans in

multiple samples. Forty-eight or sixty identical microarrays are printed on one microscope slide, segregated by hydrophobic boundaries. A set of serum samples is incubated on the arrays in a random order, and each slide is probed with a single antibody or lectin.

CHAPTER 3 GLYCAN VARIATION ON MUCINS FROM PANCREATIC CANCER

T. Yue, I.J.Goldstein, M.A.Hollingsworth, K.Kaul, R.E.Brand, B.B.Haab (2009). The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics*. 8,1697-1707

Note: Only the main displays of the above publication are provided in the dissertation and numbered as listed in “List of Tables” or “List of Figures”. All supplemental materials of this publication were numbered as in the original publication and can be found online as “Supplementary Data”.

3.1 Abstract

Changes to the glycan structures of proteins secreted by cancer cells are known to be functionally important and to have potential diagnostic value. However, an exploration of the population variation and prevalence of glycan alterations on specific proteins has been lacking due to limitations in conventional glycobiology methods. Here we report the use of a previously-developed antibody-lectin sandwich array method to characterize both the protein and glycan levels of specific mucins and CEA-related proteins captured from the sera of pancreatic cancer patients ($n = 23$) and control subjects ($n = 23$). The MUC16 protein was frequently elevated in the cancer patients (65% of the patients) but showed no glycan alterations, while the MUC1 and MUC5AC proteins were less frequently elevated (30% and 35%, respectively) and showed highly-prevalent (up to 65%) and distinct glycan alterations. The most frequent glycan elevations involved the TF antigen, fucose, and Lewis antigens. An unexpected increase in the exposure of alpha-linked mannose also was observed on MUC1 and MUC5ac, indicating possible N-glycan modifications. Because glycan alterations occurred independently from the protein levels, improved identification of the cancer samples was achieved using glycan measurements on specific proteins, relative to using the core protein measurements. The most significant elevation was the CA 19-9 antigen on MUC1, occurring in 19/23 (87%) of the cancer patients and 1/23 (4%) of the control subjects. This work gives insight into the prevalence and protein carriers of glycan alterations in pancreatic cancer and points to the potential of using glycan measurements on specific proteins for highly effective biomarkers.

3.2 Introduction

Alterations to the glycan structures on extracellular proteins are a common feature of many types of epithelial cancer such as pancreatic, colon and breast cancers [14, 16]. Cancer-associated glycan structures are thought to be functionally involved in many of the phenotypes characterizing cancer cells, including the ability to migrate, avoid apoptosis, evade immune destruction, and enter and exit the vasculature [17]. Since proteins bearing cancer-associated glycans can be shed by tumor cells into the circulation, blood-based diagnostic tests using glycan detection may be possible. A potential advantage of using glycans for diagnostics is that carbohydrate modifications of particular proteins may be altered more frequently or more specifically in certain disease states than their underlying core protein concentrations. However, in order to evaluate and employ such a strategy, the prevalence with which various structures appear and the specific proteins on which they appear must be better characterized.

Previous studies of cancer-associated glycosylation employing enzymatic, chromatographic, and mass spectrometry methods have been very effective for providing detailed information about the glycan structures produced by cancer cells, but due to the requirements for large amounts of material and the time involved to analyze each sample, these studies generally used either cell culture material or a small number of patient samples. Therefore, while many cancer-associated glycans have been identified, it is not known how often they appear, how closely they are associated with particular disease states, or the distribution of protein carriers on which they appear.

Affinity-based methods, using reagents such as lectins or glycan-binding antibodies, are a valuable complement to the above-mentioned methods. Using antibodies or lectins that bind specific glycans, one may reproducibly measure the levels of those glycans over multiple

samples. While affinity-based glycosylation studies do not provide the structural detail provided by mass spectrometry and enzymatic methods, they can provide information about the biological variation of a particular motif.

Lectins and glycan-binding antibodies have been used extensively in immunohistochemistry, for example in studies to examine the tissue distribution in pancreatic tumors of certain blood-group carbohydrates [50, 51]. Lectins have been valuable in immuno-affinity electrophoresis and blotting methods to identify cancer-associated glycan variants on major serum proteins such as α -fetoprotein [52], haptoglobin [53, 80], α -1-acid glycoprotein [55], and α -1-antitrypsin [56]. Antibodies raised against particular glycan groups, such as the Thomsen-Friedenreich antigens [71], the Lewis blood-group structures [72], and under-glycosylated MUC1 [81], also have been used to study the roles of glycans in cancer. As a means of quantifying glycans on specific proteins, lectins have been used in the capture or detection of proteins in microtiter plates [73].

We previously demonstrated an antibody-lectin sandwich array method [74] that is a valuable complement to the above methods and is ideal for profiling the prevalence of multiple glycans on multiple proteins. Glycan levels can be probed directly from biological samples, and many samples or detection conditions can be processed efficiently in a low-volume, high-throughput format [82]. This method is complementary to lectin microarrays [58, 60, 65], which are useful for measuring glycan levels on individual, purified proteins; glycan microarrays [83, 84], which are used to measure the recognition of carbohydrate structures by various glycan-binding reagents; and glycoprotein arrays [79], for examining glycosylation on proteins isolated from biological samples.

We applied this method to the study of glycan alterations on proteins in the circulation of pancreatic cancer patients. We sought to define the prevalence of various glycan alterations on particular protein carriers; and to investigate whether those measurements have advantages for cancer diagnostics relative to measurements of core proteins. We designed antibody microarrays to target members of the mucin and carcinoembryonic antigen cell adhesion molecule (CEACAM) families since some of those proteins are known to carry cancer-associated glycans. Mucins are extracellular, long-chain glycoproteins involved in the control and protection of epithelial surfaces, and the expression and glycosylation of several mucins are often altered and functionally involved in cancer [85, 86]. The CEACAM family of proteins also is functionally involved in cancer and carry cancer-associated glycans [87, 88], but the glycans on CEACAMs are less well studied than those on mucins. By measuring both glycan levels and the core protein levels of several of these molecules, we were able to investigate whether alterations to glycans can appear at a higher rate than changes to core protein abundances. The ability to test the presence of glycan structures on multiple protein carriers in multiple samples was critical to investigating these questions.

3.3 Results

3.3.1 Profiling cancer-associated glycans on specific proteins

Using a variant on standard sandwich methods to detect core protein levels (Figure 4a), the glycans on selected mucins and CEACAMs captured by antibody arrays were probed with a variety of lectins or glycan-binding antibodies (Figure 4b). The ability to print and process 48 or 60 antibody arrays on a single microscope slide enabled the efficient evaluation of multiple glycans in multiple samples (Figure 4c). A set of 46 serum samples ($n = 23$ from pancreatic

cancer patients and 23 from healthy control subjects (Supplementary Table 1) was incubated on the arrays of one microscope slide (along with two arrays incubated with TBS buffer as negative controls), and each array was probed with a detection lectin or antibody. The sample set was run 35 times (on 35 microscope slides), each time detected with one of 28 different lectins or antibodies (Figure 4c). The arrays were probed both with antibodies targeting core proteins and with lectins targeting glycans, some of which produced clearly different binding patterns (Figure 4d). Seventeen capture antibodies were used on each array (Refer to Table 1 in the original publication). The specificities of the capture antibodies for their respective targets had been confirmed by western blot and by array experiments (Supplementary Fig. 2), and dilution curves of pooled serum samples confirmed the detection of the targeted proteins in the linear response range at a 1:2 serum dilution (Supplementary Fig. 3).

These data provided an opportunity to explore the prevalence of particular cancer-associated glycan alterations on specific proteins. For each capture antibody, we examined the patterns of lectin binding among the set of serum samples, for example at the anti-MUC5AC (ab1) capture antibody (Figure 5a). Multiple detection reagents had elevated binding levels in some of the cancer patients. The glycan profiles at the anti-MUC16 capture antibody also showed multiple elevations in the cancer subjects, while the profiles at the anti-MUC1 capture antibody showed fewer elevations (Supplementary Fig. 4).

To obtain an overview of which glycans on which proteins had the greatest differences between the cancer and control subjects, we quantified the differences between the cancer and control samples for every capture antibody and every detection reagent. We used the area-under-the-curve (AUC) in receiver-operator characteristic (ROC) analysis. In ROC analysis, the sensitivity and specificity of a marker is calculated at multiple thresholds scanning the range of

values, and the AUC statistic gives a summary of the discriminating power of that marker over all thresholds. The matrix of AUC values from all the detection and capture reagents, represented as a cluster (Figure 5b), revealed that the proteins showing the most differences between cancer and control were the mucins MUC5AC, MUC16, and MUC1. The detection reagents giving the most consistently elevated discrimination were the CA 19-9 antibody and the GSL-1, AAL, PNA, and WGA lectins. Clearly divergent patterns of lectin-binding patterns were evident among the capture antibodies. On the basis of this analysis, we focused subsequent analyses on the mucin proteins and their glycans.

The validity of these cancer-associated patterns in glycan expression is supported by the fact that each set of samples was run independently, using several different batches of microarrays, each with independently-randomized samples. Repeat analyses using several of the detection reagents showed that the expression patterns were highly reproducible (Supplementary Fig. 5 and Supplementary Table 2). In addition, sugar competition assays confirmed that the lectin binding was specific to particular glycans (Supplementary Fig. 6 and Supplementary Table 3). Furthermore, immunoprecipitations of MUC1 followed by Western blot detection of the particular glycan levels confirmed the accuracy of the microarray measurements (Supplementary Fig. 7). The glycan levels were not associated with the age or gender of the subjects ($p > 0.05$), as determined by examining the correlations between the glycan levels and age within each patient class, or by performing a t-test on the values grouped by gender (Supplementary Table 6), which indicates that these particular glycan alterations are more likely associated with cancer than with demographic or clinical factors.

3.3.2 Prevalence of glycan changes relative to core protein changes

The ability to obtain both protein and glycan measurements at the same capture antibodies (Figure 4a-b) enabled an exploration of the relationships between these levels. In particular, for the glycan measurements with altered levels in the cancer samples, we sought to determine whether the glycan structures were changing relative to the core protein levels or simply along with the protein levels. We calculated glycan:protein ratios by dividing each glycan measurement by its corresponding core protein measurement, and we compared the glycan:protein ratios between the cancer samples and the control samples. The core protein levels were measured using array-based sandwich assays (as depicted in Figure 4a) for the mucins MUC1, MUC5AC, and MUC16 (Figure 6a). All three had elevated levels in cancer ($p = 0.03$, 0.008 , and 0.001 for MUC5AC, MUC16, and MUC1, respectively, Mann-Whitney rank-sum test). The glycan levels, measured using detection by the jacalin lectin at each capture antibody (as depicted in Figure 4b), were significantly elevated only on MUC5AC and MUC16 (Figure 6b). In contrast, the glycan:protein ratios were elevated only for MUC5AC (Figure 6c). This result indicates that the glycan target of jacalin was only elevated on MUC5AC, although all three core proteins showed cancer-associated elevations. The prevalence of the elevations, relative to healthy individuals, was estimated using a threshold set to the level of the second-highest control subject (one false positive out of 23, or 96% specificity). The prevalence of glycan elevations on MUC5AC (65%, Figure 6b) was higher than the prevalence of MUC5AC protein elevations (35%, Figure 6a), suggesting that some patients had glycan elevations on MUC5AC independent of protein elevations.

Similar comparisons were made for all the glycan:protein ratios. The AUC for the discrimination of cancer from control was calculated using each glycan:protein ratio (Figure 6d)

such that a high AUC indicates a cancer-associated elevation in the glycan relative to the protein. Certain ratios had AUCs near or above 0.8 and had significant ($p < 0.05$, Mann-Whitney test) elevations in cancer. The most prevalent elevation was the CA 19-9 antigen on MUC1 (15/23 patients, or 65%), although MUC5AC showed the greatest number of different cancer-associated elevations. The pattern of cancer-associated glycan alterations on MUC5AC was clearly different than both MUC1 and MUC16. MUC1 shared elevations in the CA 19-9 antigen and the glycan target of GSL-I, but was missing elevations in the glycan targets of others, such as Jacalin and AAL. MUC16, in contrast, showed no significant elevations in glycan:protein ratios.

3.3.3 Discrimination of cancer from control using glycan detection

Next we examined whether it was possible to achieve more accurate discrimination of cancer from control using glycan measurements, relative to using protein measurements. This question relates to whether glycan and protein elevations occur independently or together in the same patients. If glycan and protein elevations occur together in the same patients, minimal additional discrimination of cancer from control would be achieved using glycan detection. A comparison of the CA 19-9 antigen on MUC5AC with the MUC5AC protein levels shows that 10 (45%) of the patients had glycan elevations without protein elevations (Figure 7a). Of the 10 patients with elevations in the glycan:protein ratio, four also had protein elevations. Therefore, for this glycan on MUC5AC, elevation can occur independently of protein elevation but also coincide with protein elevation. A similar relationship was observed for the CA 19-9 antigen on MUC1 (Figure 7a). For MUC16, the strong correlation between the glycan level and the protein level indicates that the CA 19-9 antigen is rarely altered in cancer relative to the core protein level (Figure 7a).

This result indicates that, for certain proteins and glycans, the measurement of glycans on proteins could provide better cancer detection than just measuring the core protein levels. ROC curves comparing protein detection to glycan detection show that, in cases where the glycan can be upregulated independently of the protein, better discrimination of cancer from control is achieved using glycan measurements (Figure 7b). At a threshold of 96% specificity (one false positive out of 23 control subjects), 18 of 23 (78%) and 19 of 23 (83%) cancer subjects showed elevations of CA 19-9 on MUC5AC and MUC1, respectively. For MUC16, since glycans are not elevated when the protein is not elevated, no improvement in discrimination is observed.

3.3.4 Structural insights from lectin-binding profiles

Knowledge about the binding specificities of each lectin can provide insights into the similarities and differences between the normal and cancer-associated glycan structures. For each capture antibody, we organized the detection lectins and antibodies according to their primary specificities, and we examined the overall discrimination of cancer from control using both the total glycan level and the glycan:protein ratio (Figure 8). An elevated glycan level indicates that the indicted glycan is present on the protein in the cancer patients, but not necessarily elevated relative to the core protein level, and an elevated glycan:protein ratio indicates a particular glycan was elevated relative to the core protein. Notable differences and similarities were observed among the three mucins. The most consistent elevation in glycan:protein ratio was in fucose, with even MUC16 showing elevations using the lectin AAL. MUC5ac showed evidence of elevations in Gal-GalNAc, which forms the O-glycan core-1 structure, known as the TF antigen. MUC1 displayed the TF glycan in cancer, but not strong elevations relative to the protein levels. Gal-GlcNAc disaccharides, which are characteristic of extended chains on both N-glycans and O-glycans, were found most strongly on MUC16 in cancer, but may have been

elevated most on MUC1. Terminal GalNAc, characteristic of truncated O-glycans displaying the Tn antigen, was present but not strongly elevated in cancer relative to the core protein, as was terminal and poly-GlcNAc. Terminal mannose was found on all three mucins and elevated relative to the protein level on MUC1 and MUC5AC. This result was unexpected because O-glycosylation, which does not contain mannose, is the main type of glycosylation found on mucins, and cancer-associated alterations to N-glycans on mucins were not previously recognized.

3.4 Discussion

Despite repeated observations of cancer-associated modifications to structures of carbohydrate chains, the prevalence of specific glycan alterations, and the relationships to their carrier proteins, has not been characterized. Such information would be useful for determining potential involvement in cancer processes or usefulness for cancer detection. We have shown here the profiling of selected glycans and carrier proteins using a novel antibody-glycan sandwich array method, resulting in the characterization of the rate of alterations to particular glycans on MUC1, MUC5AC, and MUC16.

The results revealed significantly different behaviors between MUC1, MUC5AC and MUC16. MUC16 was the most frequently elevated at the core protein level (in 65% of the patients), but it showed few glycan alterations. MUC5AC was elevated at the protein level (in 35% of the patients) and had the most glycan alterations, while MUC1 was weakly elevated and showed a few highly prevalent glycan alterations. The clear differences between the mucins in their regulation of glycosylation suggest distinct functions for these mucins and their glycans in cancer.

By running the assay either for protein detection or glycan detection, we were able to explore the relationships between those levels. The glycan alterations on MUC1 and MUC5AC occurred independently of protein elevations, so that some patients showed glycan elevations without protein elevations and other patients showed the opposite trend. The most prevalent elevation in glycan:protein ratio, the CA 19-9 antigen on MUC1, occurred in 65% of the patients, compared to a largely non-overlapping 35% of the patients with elevations in MUC1 core protein. These two behaviors might represent differential responses of cancers, either releasing greater amounts of mucins without altering the glycans, or instead releasing less mucin but increasing a particular glycan epitope.

Because of the complementary relationship between protein and glycan elevations, measuring the glycan on the specific protein resulted in detecting a higher percentage of patients (83% for CA 19-9 on MUC1) relative to measuring just the protein. This fact points to the possible future usefulness of similar assays for cancer detection, as suggested before in studies of fucosylated alpha-fetoprotein [89] and haptoglobin [53, 54] for identifying liver and pancreatic cancers. Validation of potential markers will involve studies on larger, blinded sample sets including samples from conditions that might potentially give elevations, such as benign liver diseases and inflammatory states of the pancreas. In addition, it will be important to develop protein standards for potential markers. Standards are useful for calibrating the assays to achieve quantitative, absolute measurements and for comparing results between platforms and laboratories. The development of standards for glycan measurements on particular proteins is challenging since both the glycan and the protein must be well characterized, so that task likely will be undertaken once the most valuable biomarkers have been identified and their glycan structures characterized. Future biomarker work may focus on glycan alterations of MUC5AC,

which was not previously recognized as a serological marker for pancreatic cancer, perhaps because the protein level alone did not provide good cancer detection. Since MUC5AC is not expressed in the normal pancreas but shows expression in pre-malignant pancreatic intraepithelial neoplasia (PanIN) lesions [90, 91], it could have value for early detection if glycan structures can be found that are unique to incipient cancer.

Knowledge of the specificities of the lectins and detection antibodies provided insights into the nature of the structural alterations associated with cancer. The most frequent elevations involved fucose, the CA 19-9 antigen, the TF antigen (Gal β 1,3GalNAc), and terminal mannose. The elevation of the TF antigen relates to previous observations of truncated O-glycans on MUC1 expressed by cancer cells, resulting in the exposure of the TF and Tn (GalNAc-O-Ser/Thr) antigens [14, 85, 92-96]. The exposure of these antigens has been found in the tissue of about 90% of all carcinomas [97], while we found it on MUC5AC in 65% of the cancer patients (using Jacalin detection). The prevalence in blood may be less than that in tissue due to incomplete release into the circulation. Our results suggested that both the TF and Tn antigens were present in cancer on all three mucins, but just the TF antigen was elevated relative to the core protein level. A dominance of the TF over the Tn antigen in pancreatic cancer, which was not previously noted, could indicate functional importance of the TF antigen. The increased exposure of the TF antigen can lead to increased stimulation through the c-Met or MAPK pathways [98] by endogenous lectins or lectins of bacterial or nutritional origin, as in colon cancer [99, 100]. Galectin-3 is an endogenous lectin that can bind TF on MUC and enhance the adhesion of cancer cells to endothelial cells, possibly by the clustering of MUC1 to reveal underlying adhesion molecules [101]. The increased exposure of the TF antigen on MUC5AC was not previously observed and could indicate a broader presentation of that epitope than previously appreciated.

The increased binding of the CA 19-9 antibody indicates increased levels of the sialyl-Lewis A carbohydrate antigen on MUC1 and MUC5AC. The Lewis structures are blood group carbohydrate antigens (antigens found on red blood cells whose structures are determined by genetic variants) on the termini of both O- and N-glycans. Elevated sialylated Lewis epitopes on mucins were observed previously on MUC1 [74, 94, 102], and we found here elevations also on MUC5AC. The Lewis antigens are ligands for selectin receptors on lymphocytes and endothelial cells [92], so the increased presence of Lewis antigens on the surfaces and secretions of cancer cells has implications for the ability of cancer cells enter and exit the vasculature [103].

Since the Lewis antigens contain α 1,3- or α 1,4-linked fucose, their elevation could relate to the elevated binding of the lectin AAL, which targets fucose and which showed strong binding on all three mucins. Another lectin that targets fucose, UEA, is specific to α 1,2-linked fucose and was only elevated on MUC1 and MUC5AC. α 1,2-linked fucose is typically found at the core GlcNAc of N-glycans, suggesting alterations to N-glycans on MUC1 and MUC5AC. Generally increased levels of fucose groups, independent of Lewis antigens, have been seen in a variety of conditions, particularly on the protein haptoglobin [80].

An unexpected result was the elevated binding of the lectins ConA and LCA on MUC1 and MUC5AC. The mannose structures targeted by these lectins are typically found on N-glycans, which are present on mucins [104] but were not known to have cancer-association alterations on mucins. This possibility of altered N-glycans is consistent with the observation noted above of altered 1,2-linked fucose. Therefore, N-glycans as well as O-glycans may play roles in modifying the behaviors of MUC1 and MUC5AC in disease. Recently, high-mannose structures were found to be immunogenic cancer antigens on melanoma cells [105]. Future

studies could explore the functions of these various glycan structures found on secreted mucins in cancer patients.

Lectin-binding patterns do not give precise information on glycan structures, so it will be necessary in future studies to relate these data to results from other structural analysis methods, including mass-spectrometry analyses or complementary technologies such as lectin arrays [58] or glycoprotein arrays [79]. It also will be valuable to determine whether increases in certain structures are due to an increased number of glycan sites being glycosylated, a possibility suggested by a previous study of MUC1 [106], or a shift in structures on the same number of sites.

Several factors might contribute to the glycan alteration or affect the lectin binding. Variation exists between individuals in native glycosyltransferases activities or abilities to elaborate certain structures, for example the blood group antigens on red blood cells. Certain individuals have genetic variants that prevent the addition of α 1,3- or α 1,4-linked fucose on secreted proteins and thus would not be expected to produce elevated blood levels of the Lewis antigens [107]. The individuals in this study that were negative for CA 19-9 elevations may represent those genotypes. Other glycans potentially could be used to complement CA 19-9, a strategy previously attempted using various related monoclonal antibodies [108]. However, only limited detection of CA 19-9-negative patients was achieved. Similarly, we did not find any glycans that significantly complemented CA 19-9, as certain patients (4 out of 23, 17%) were negative for all markers tested. An important research goal will be to improve the sensitivity of diagnostic assays by identifying markers that are complementary to CA 19-9, while achieving high specificity.

Another factor potentially influencing the assay is the fact that the capture antibody affinities might be affected by glycosylation status, as observed for several MUC1 antibodies [109]. Among the MUC1 antibodies used here, the HMGF-1 clone is insensitive to the addition of a GalNAc or a Gal β 1,3GalNAc to its peptide epitope [110]; the SM3 clone has a preference for an epitope with short, non-branched O-glycan chains (known as the “core 1” structure) [111]; and the sensitivity to glycosylation of the CA 15-3 antibody used for much of the analysis is not known. The main MUC5AC antibody used here (the 45M1 clone) binds an epitope in one of the cysteine-rich domains of MUC5AC [112], which is not heavily glycosylated, although antibody binding is affected by reduction [112], indicating the importance of the secondary structure. This diversity in antibody behavior highlights the value of testing multiple antibodies in a multiplexed setting. It also shows the need to regard each antibody as is its own assay which measures a subset of all molecules of that type. It will be important to characterize the isoforms and glycan structures bound by particular antibodies in order to better determine which structures are most strongly associated with cancer.

In summary, the application of this new method to the study of sera from pancreatic cancer patients resulted in the characterization of the prevalence and nature of certain glycan alterations and their potential use for pancreatic cancer diagnostics. We identified clear differences between the mucins, with MUC16 showing the strongest protein elevation but no glycan alterations, MUC5AC showing the most glycan alterations, and MUC1 showing slight protein elevations with selected, highly-prevalent glycan alterations. The most prevalent glycan alterations were fucose, Lewis structures, the TF antigen, and, unexpectedly, terminal mannose. Because certain glycans on MUC1 and MUC5AC were altered independently of core protein elevations, the measurement of those glycan on the core proteins resulted in improved cancer

detection relative to the measurement of core proteins alone. Future studies will build upon these findings to further characterize glycan alterations in pancreatic cancer and develop their uses for patient care. This technological approach should be useful for additional studies in biomarker discovery or basic glycobiology.

3.5 Materials and Methods

Serum samples. Serum samples from pancreatic cancer and healthy subjects were collected at Evanston Northwestern Healthcare under protocols approved by a local Institutional Review Board. The control samples were collected from high-risk individuals from pancreatic-cancer-prone families undergoing surveillance with Endoscopic Ultrasound (EUS) or Endoscopic Retrograde Cholangiopancreatography (ERCP). The control subjects had no pancreatic lesions. A summary of the patient characteristics is given in Supplementary Table 1. All samples were stored at -80°C and sent frozen on dry ice. Each aliquot had been thawed no more than three times before use.

Antibodies and lectins. The antibodies, lectins, and carbohydrates were obtained from various sources (see Supplemental Tables 2 and 3). All antibodies were screened for reactivity and integrity, purified, and prepared at 0.5 mg/ml in pH 7.2 phosphate-buffered saline (PBS). See Supplementary Figure 1 for information on the antibody preparation steps.

Microarray fabrication. Approximate 350 pl of each antibody solution was spotted on the surfaces of ultra-thin nitrocellulose-coated microscope slides (PATH slides; GenTel Biosciences) by a piezoelectric non-contact printer (Biochip Arrayer; PerkinElmer Life Sciences). Forty-eight identical arrays containing triplicates of all selected antibodies were printed on each

slide. Hydrophobic borders were imprinted around each array using a stamping device [82] (SlideImprinter, The Gel Company, San Francisco, CA).

Microarray assays. Microarray assays were performed to measure either protein levels (Figure 4a) or the glycan levels (Figure 4b) on the captured proteins. Protein levels were determined using a dual-antibody sandwich assay (Figure 4a). The sandwich assay consisted of four 1-hour-incubations in room temperature (RT) with the following reagents: 1) blocking buffer (PBS containing 0.5% Tween-20 (PBST0.5) and 1% BSA); 2) a serum sample, diluted two-fold in 1XTBS containing 0.08% Brij, 0.08 Tween-20 and 50 µg/ml protease inhibitor cocktail (Complete Protease Inhibitor Tablet, Roche Applied Science); 3) biotinylated detection antibody (2 µg/ml), diluted in PBST0.1 containing 0.1% BSA; 4) streptavidin-phycoerythrin (10 µg/ml, Roche Applied Science), diluted in PBST0.1 containing 0.1% BSA. After each step, the slides were rinsed in three baths of PBST0.1 and dried by centrifugation (Eppendorf 5810R, rotor A-4-62, 1500 x g).

The measurement of glycans on the captured proteins (Figure 4b) was carried out as above, except the glycans on the spotted antibodies were derivatized [74] prior to use (see Supplementary Methods), and the arrays were probed with glycan-binding lectins or antibodies.

Fluorescence emission from the phycoerythrin was detected at 570 nm using a microarray scanner (LS Reloaded, Tecan). All arrays within one slide were scanned at a single laser power and detector gain setting. The images were quantified using the software program GenePix Pro 5.0 (Molecular Devices, Sunnyvale, CA). Spots were identified using automated spot-finding or manual adjustments for occasional irregularities. The local backgrounds were subtracted from the median intensity of each spot, and triplicate spots were averaged using the geometric mean. The coefficient of variation between replicate analyzed spots was 10-15%.

Statistical analysis. The area-under-the-curves (AUC) were calculated from the receiver-operator characteristic (ROC) analysis using a custom script. Pearson correlations and Student's T-tests were calculated using Microsoft Excel. The Mann-Whitney rank-sum tests were performed using OriginPro 8. Clustering and visualization were performed using the programs Cluster and Treeview and MultiExperiment Viewer.

Additional methods. Details of the immunoprecipitation, western blot, sugar competition, and antibody derivatization procedures are available in the supplementary materials.

APPENDICES

FIGURES

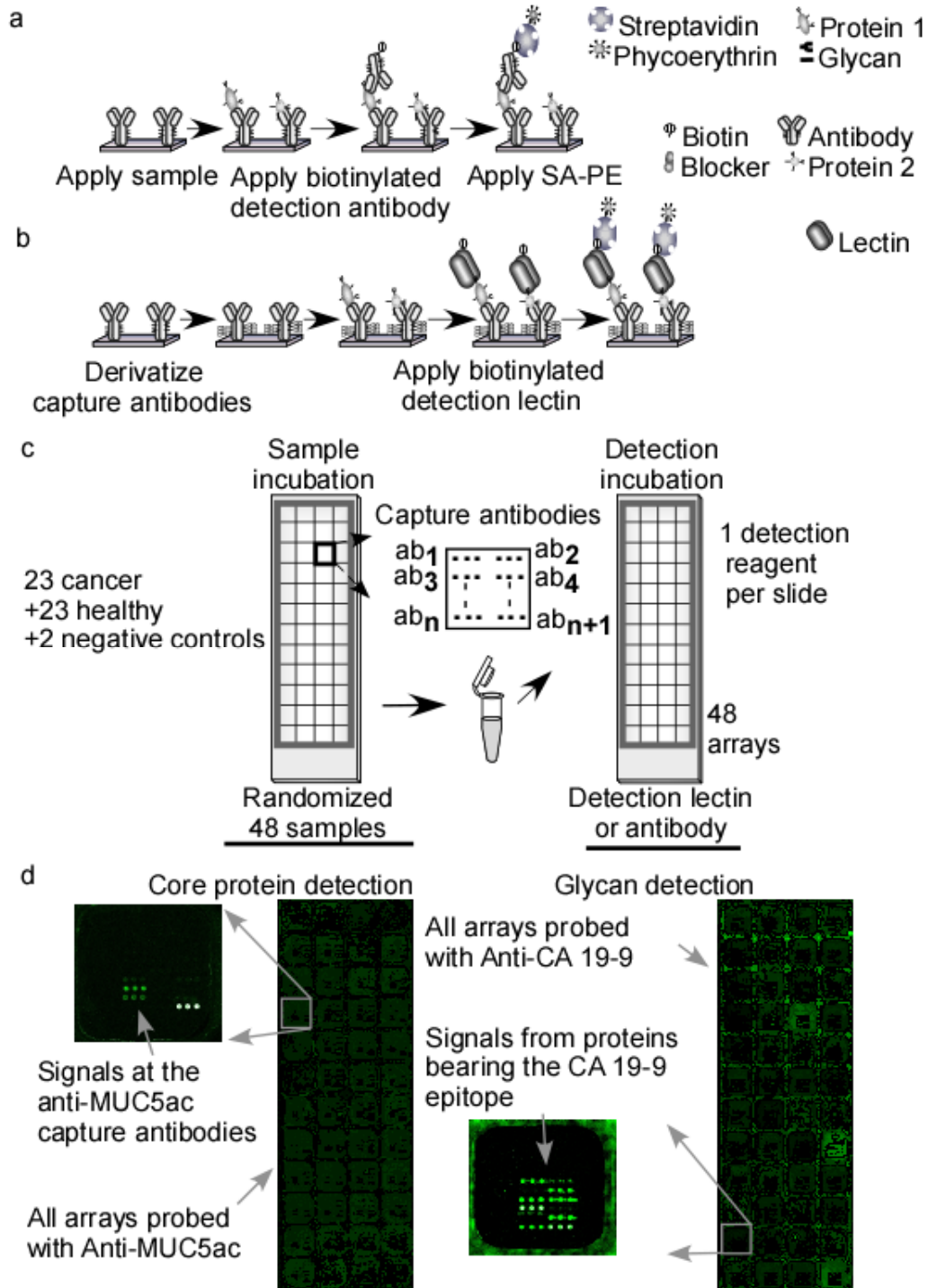


Figure 4 Protein and glycan detection on antibody arrays.

a) Array-based sandwich assays for protein detection. Multiple antibodies are immobilized on a planar support, and the captured proteins are probed using biotinylated

detection antibodies, followed by fluorescence detection using phycoerythrin-labeled streptavidin. b) Glycan detection on antibody arrays. This format is similar to above, but the detection reagents target the glycans on the capture proteins rather than the core proteins. The glycans on the immobilized antibodies are chemically derivatized to prevent lectin binding to those glycans. c) High-throughput sample processing. Forty-eight or sixty identical microarrays are printed on one microscope slide, segregated by hydrophobic boundaries. A set of serum samples is incubated on the arrays in a random order, and each slide is probed with a single antibody or lectin. d) Example antibody array results for core protein detection (left) and glycan measurement (right). SA-PE, streptavidin-phycoerythrin.

For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

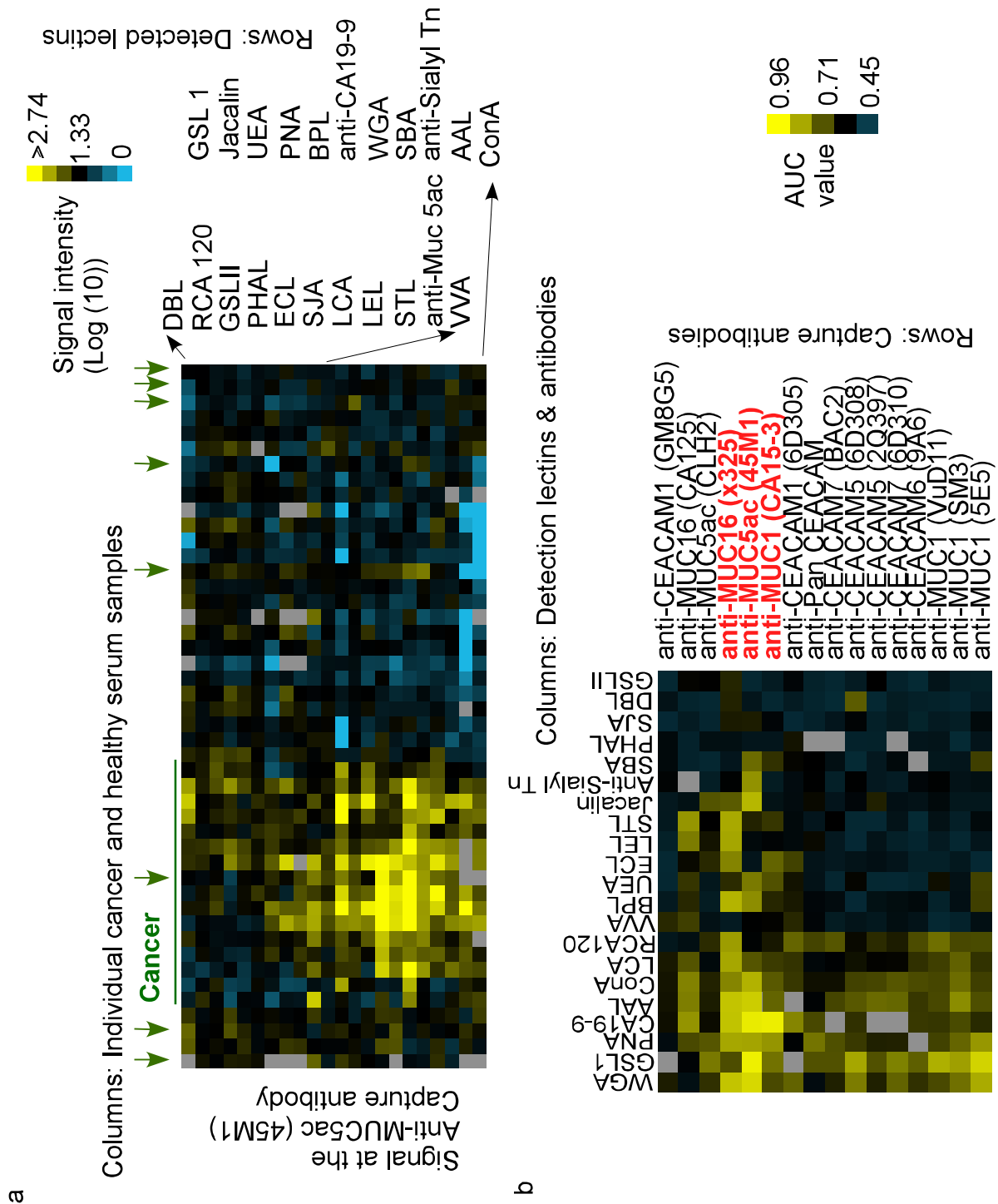


Figure 5 Cancer-associations of glycan levels.

a) Glycan patterns at the anti-MUC5ac capture antibody. The signal levels at the anti-MUC5ac (clone 45M1) capture antibody, for each sample and each detection reagent, were

clustered by similarity. The values were log-transformed (base 10) and median-centered along the row axis. The log-transformed intensities are represented by the color bar. C, cancer; N, normal. b) For each detection reagent and each capture antibody, the signal levels were compared between the cancer and control serum samples, resulting in an AUC value. The matrix of AUC values was clustered by similarity, and the AUC values are represented by the color bar. Gray boxes indicate missing data.

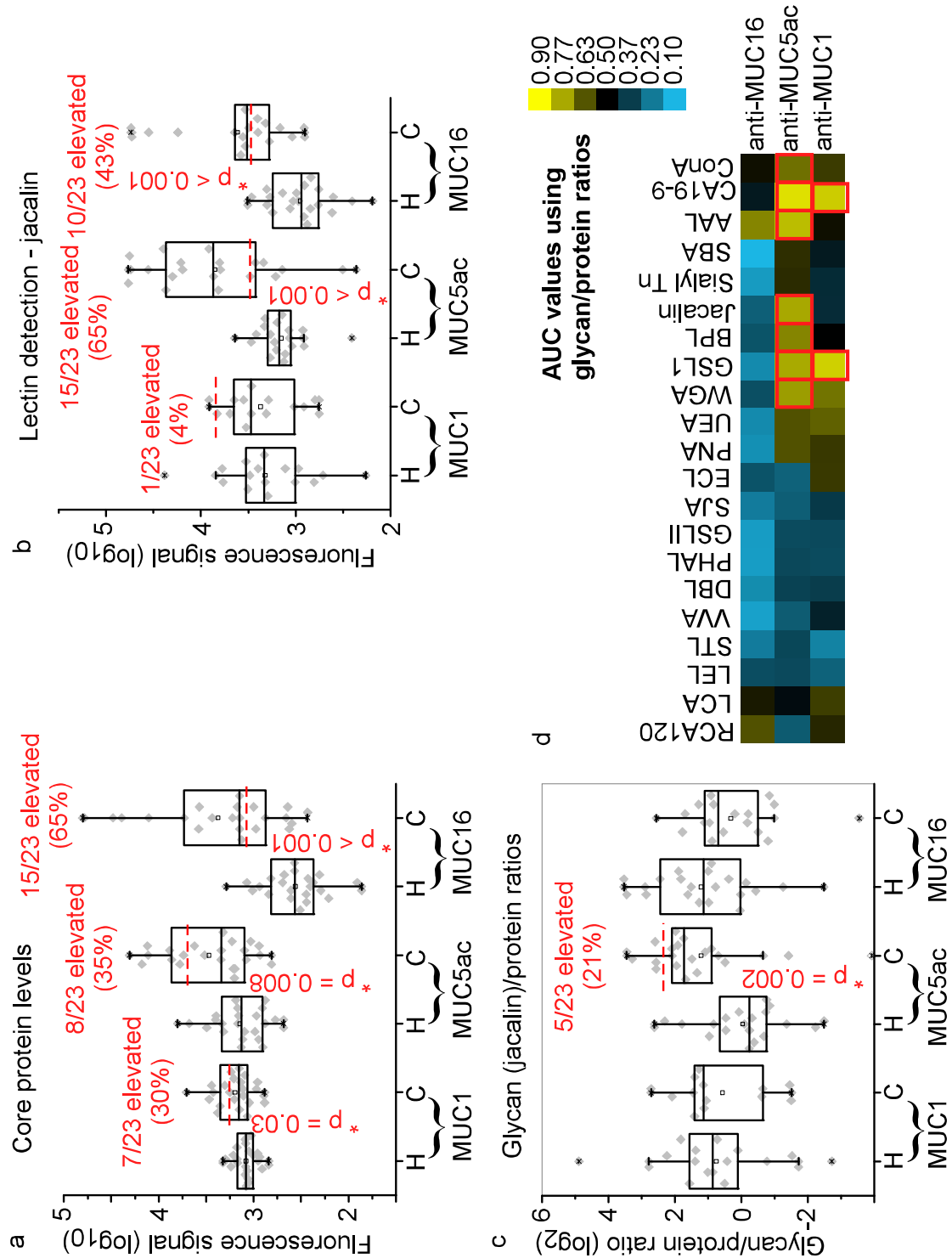


Figure 6 Glycan elevations relative to core protein elevations.

a) Core protein levels. Using an antibody sandwich assay (Figure 4a), the protein levels of MUC1 (Ab CA 15-3 for capture and detection), MUC5AC (Ab 45M1 for capture and

detection), and MUC16 (Ab x325 for capture and CA 125 for detection) were measured. Each gray diamond represents a sample from the healthy controls (H) or cancer patients (C). The box gives the upper and lower quartiles, the vertical lines give the ranges, and the horizontal line gives the median. The p-value was calculated using the Mann-Whitney test, and the threshold for elevation in the cancer patients (dashed red line) is based on the second-highest control subject. b) Glycan levels at the same capture antibodies as in panel (a) using detection with the Jacalin lectin. c) Glycan:protein ratios. The glycan level for each sample in panel (b) was divided by the protein level of the corresponding sample and capture antibody in panel (a). d) AUCs using the glycan:protein ratios. The AUC for discriminating cancer from control was calculated for each set of glycan:protein ratios, and the combined AUCs were hiererarchically clustered.

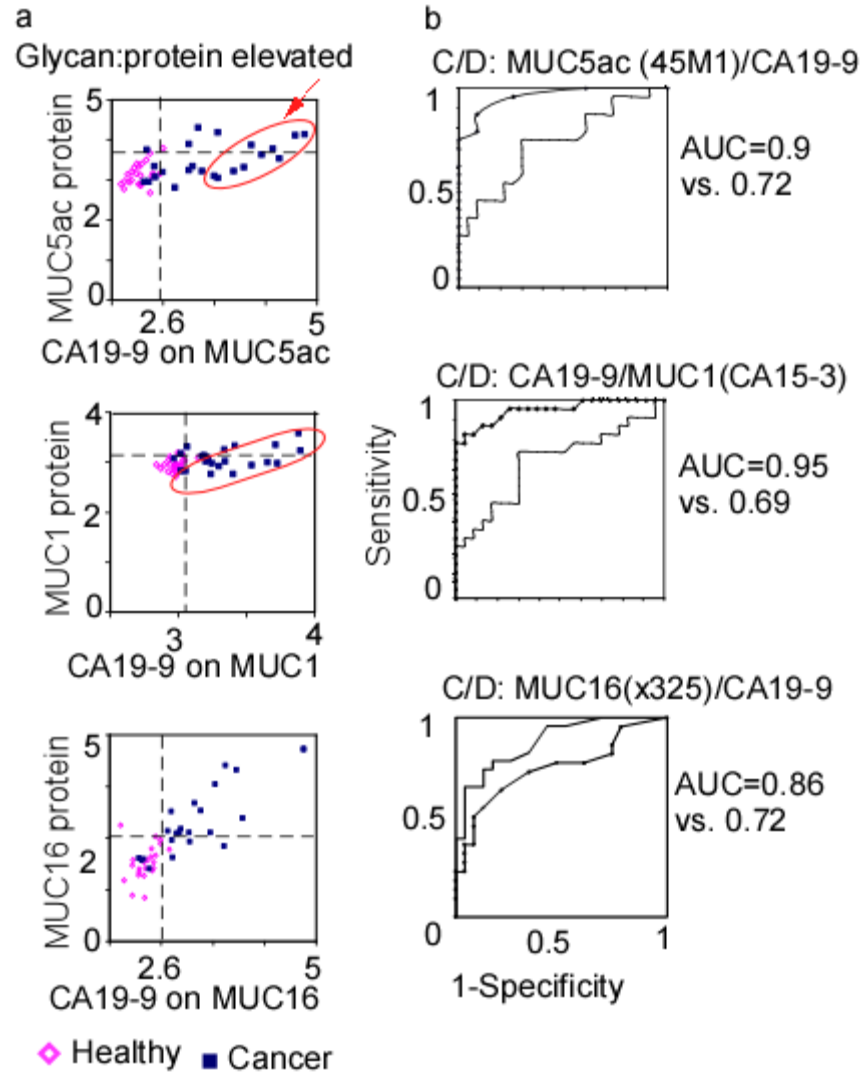


Figure 7 Discriminating cancer from control using glycan or protein measurements.

a) Correlations between protein and glycan levels. Glycan levels are plotted on the horizontal axes, and corresponding protein levels are plotted on the vertical axes, for the indicated proteins and glycans. The dashed lines represent thresholds based on the second-highest control sample in each dimension. Each point represents a sample from the indicated patient group. The red ovals indicate samples with elevated glycan:protein ratios, determined using the same calculation as in Figure 6c. The values were log-transformed (base 10). b) Comparison of biomarker performance using protein and glycan levels. Receiver-operator

characteristic curves were generated using either the protein measurements (smooth curve) or a glycan measurement on the corresponding protein (solid diamonds), for the indicated capture antibodies and detection reagents. C, capture; D, detection.

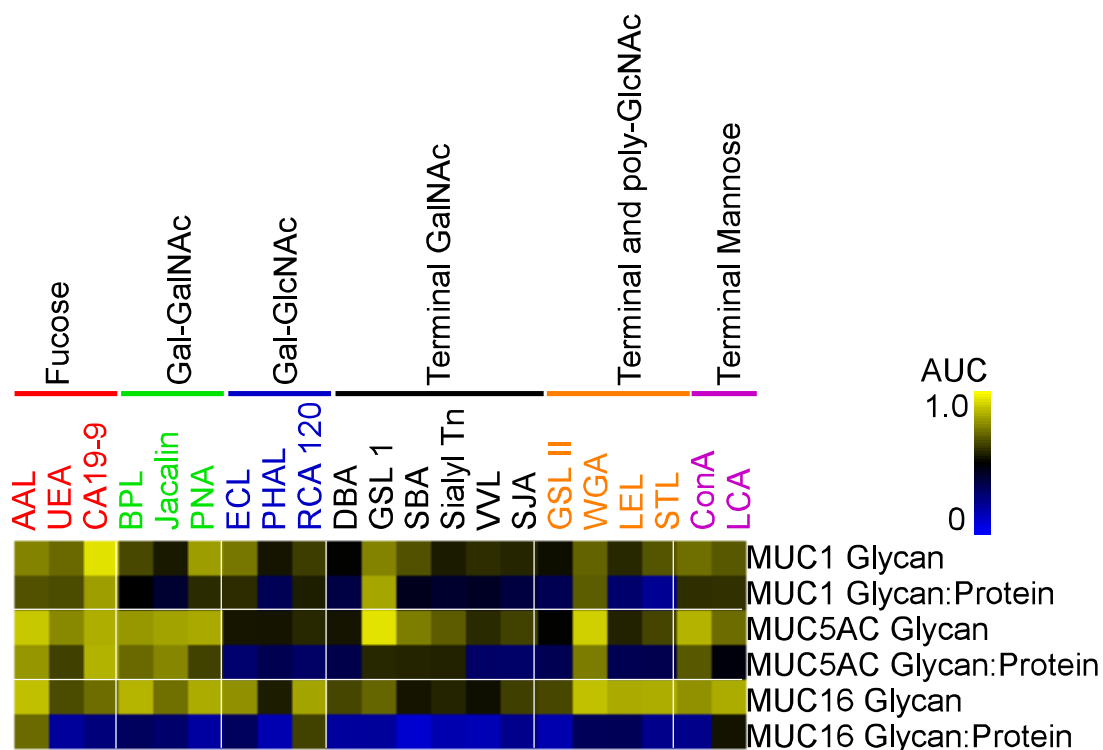


Figure 8 Cancer-associations of major structural features.

The AUC for discriminating cancer from control calculated using either glycan measurements or glycan:protein ratios are presented for each mucin and each detection reagent. The detection reagents are organized according to primary specificities, although some specificities are more complex or diverse than indicated here.

CHAPTER 4 CA19-9 ON SPECIFIC PROTEINS

T. Yue, L. Li, M.A.Anderson, D.E.Brenner, D.M.Simeone, Z.Feng, R.E. Brand, B.B.Haab.
Enhanced sensitivity of pancreatic cancer detection by measuring the CA19-9 antigen on specific protein carriers. (In preparation for submission to *JNCI*)

4.1 Abstract

The goal of this study was to develop a blood test that can detect a higher percentage of pancreatic cancer patients than the tumor marker CA 19-9. We tested the hypothesis that the measurement of the CA 19-9 antigen (an oligosaccharide found on multiple proteins) on individual proteins could contribute to improved performance over the standard CA 19-9 assay, which measures the CA 19-9 antigen on all proteins. Serum or plasma samples were incubated on microarrays containing antibodies against the mucin proteins MUC1, MUC5AC, and MUC16. After the proteins were captured by the immobilized antibodies, the levels of the CA 19-9 antigen on the captured proteins were measured by incubation of the CA 19-9 monoclonal antibody. Four sample sets from three different institutions were examined, with a total of 333 individual samples from patients with pancreatic adenocarcinoma or pancreatitis. The CA 19-9 marker distinguished cancer from benign disease with 84-91% sensitivity at 75% specificity. The measurements of CA 19-9 on individual protein carriers contributed complementary information to the standard CA 19-9 marker. Thresholds could be set to detect elevations in many of the patients that showed no CA 19-9 elevations, while not increasing the false positive rate. In all sample sets, the panel showed improved sensitivity (85%-100%) over total CA19-9 alone (79%-90%). Among all false negatives classified by total CA19-9, the additional three markers in the panel picked up from 25% to 100% of them. The consistent performance over independent sample sets supports the generality and reliability of the novel marker panel as a tool for improving the accuracy of diagnoses of pancreatic cancer.

4.2 Introduction

Several factors contribute to the extremely poor prognosis associated with pancreatic cancer, including the resistance of the disease to available therapeutic options, its tendency to metastasize at small primary tumor sizes, and its induction of systemic metabolic problems [113]. The lack of effective tools for accurately detecting and diagnosing the disease at early stages further contributes to the problems in treating the disease. Because of the lack of early detection methods, most pancreatic cancers are detected at an advanced stage. Furthermore, because established disease can be difficult to diagnose due to clinical similarities with certain benign diseases such as pancreatitis [114], some patients may receive sub-optimal treatment. The current diagnostic modalities include non-invasive imaging, endoscopic ultrasound, and cytology based on fine-needle aspiration [115]. These methods are useful for identifying pancreatic abnormalities and rendering an accurate diagnosis in many cases, but they come with high cost, significant expertise required for interpretation, and inherent uncertainty. Molecular markers could provide a useful complement to imaging and cytology methods, since they have the potential to provide objective information in an inexpensive, routine assay. Therefore, identifying and developing molecular markers providing useful diagnostic information for pancreatic cancer is a high priority.

The CA 19-9 serum marker is elevated in the majority of pancreatic cancer patients but does not achieve the performance required for either early detection or diagnosis, due to both false positive and false negative readings [116]. Patients with biliary obstruction, liver diseases, and pancreatitis may have elevations in CA 19-9, so it is not exclusively specific for malignancy. In addition, some patients with cancer do not show elevations [107], reducing its usefulness for

confirming cancer in suspect cases. The information from CA 19-9 is useful, in coordination with other clinical factors, for monitoring disease progression in patients receiving therapy [117].

The CA 19-9 marker is a carbohydrate antigen that is detected by a monoclonal antibody. This carbohydrate antigen, a quatra-saccharide called sialy Lewis A, is found on multiple, different proteins. In the sandwich ELISA method used in the clinical assay, the CA 19-9 monoclonal antibody measures the CA 19-9 antigen on many different carrier proteins [118]. The identities of the carrier proteins are not well characterized but are known to include mucins and carcinoembryonic antigen [102, 118]. It is possible that the carrier proteins of the CA 19-9 antigen are different between disease states, as suggested earlier [85]. If that is the case, the detection of the CA 19-9 antigen on particular carrier proteins may yield improved discrimination of the disease states, in comparison to measurements of total CA 19-9.

We previously demonstrated a method for detecting the level of particular glycans on individual proteins captured out of biological solutions [74, 119, 120]. Antibody arrays capture multiple, different proteins, and glycan-binding lectins or antibodies detect the glycan levels on the captured proteins. This method provides sensitive and reproducible measurements in low sample volumes and is compatible with high-throughput sample processing [82]. The antibody-lectin sandwich array is ideal for measuring the levels of the CA 19-9 antigen on multiple, specific proteins in multiple samples. Previous work showed that the mucins MUC1, MUC5AC, and MUC16 are major cancer-associated carriers of the CA 19-9 antigen in the blood [119]. In this work, we tested the hypothesis that the detection of the CA 19-9 antigen on specific proteins can yield improved biomarker performance over the detection of total CA 19-9. We tested this hypothesis for the particularly difficult diagnostic problem of differentiating pancreatic cancer patients from pancreatitis patients [114], for which CA 19-9 alone does not give sufficient

performance to be clinically useful. We show that subgroups of cancer patients are evident based on their CA 19-9 carrier proteins, and that a biomarker panel based on the detection of the CA 19-9 on specific proteins identifies a greater percentage of cancer patients than the conventional CA 19-9 assay.

4.3 Results

4.3.1 Profiling cancer-specific glycans on specific proteins

Antibody arrays were generated to target CA 19-9 and the mucin proteins MUC1, MUC5AC, and MUC16. The mucins MUC1, MUC5AC, and MUC16 were targeted based on results from a previous study [119]. Four to five different monoclonal antibodies were used for each protein, and each antibody was printed in triplicate. The locations of the triplicate spots were randomized to minimize potential positional bias within each array. Serum samples were incubated on the arrays, and the arrays were probed with the CA 19-9 antibody to detect either the total level of its target antigen (detected at the CA 19-9 capture antibody) or its level on particular proteins (detected at the capture antibodies against specific proteins) (Figure 9). The ability to print and process 48 antibody arrays on a single microscopic slide enabled the efficient evaluation of multiple clinical samples (Figure 9a). Dilution curves of pooled serum samples generated in our previous study [119] confirmed the detection of the targeted proteins or glycans in the linear response range at a 1:2 serum dilution, and the use of negative control antibodies (mouse mAbs lacking specificity for any human protein) and negative control arrays (arrays incubated with PBS buffer instead of serum) confirmed a lack of non-specific binding to the capture antibodies by the detection reagents. The various capture antibodies displayed distinct binding patterns (Figure 9b) consistent with the unique specificities of the antibodies.

Three independent sample sets, obtained from three different institutions, were processed (Table 1). Sample set #3 was processed blinded and in duplicate on different days with distinct batches of microarrays. As in initial comparison of total CA 19-9 to CA 19-9 on individual proteins, we characterized the performance of the individual measurements for discriminating between pancreatic cancer and pancreatitis. In each set, the total CA 19-9 level was significantly higher in the cancer patients than in the pancreatitis patients ($p < 0.001$, student's t-test). The sensitivity of detecting cancer was 79-90% at a specificity of 75% in the three sets (Figure 10), which included discrimination of both early-stage and late-stage pancreatic cancer patients from pancreatitis patients. The performance of CA 19-9 detection on the individual mucins was similar to or slightly better than total CA 19-9 for certain antibodies in each set (Table 2). The fact that similar discrimination could be achieved between total CA 19-9 and CA 19-9 on individual carriers indicates that these mucin proteins are major disease-associated carriers of the CA 19-9 antigen. However, no individual marker showed a consistent, significant improvement over total CA 19-9.

4.3.2 Investigation of a panel to improve detection sensitivity over CA 19-9

Although no individual marker significantly out-performed total CA 19-9, complementarity between the measurements might yield improved performance if used in combination. The antibody array platform enables efficient investigations of this question. The potential for improved biomarker performance using a panel of markers was supported by the lack of correlation among the measurements. The total CA 19-9 measurements did not significantly correlate with CA 19-9 on any individual proteins, nor did measurements correlate between the individual proteins.

Each of the three datasets revealed a subgroup of about 15% of the cancer patients which showed no elevation in total CA 19-9 (Figure 10). Therefore we asked whether particular measurements are complementary to CA 19-9 by enabling the detection of this subgroup of patients. Such measurements could provide improvements in sensitivity and in overall biomarker performance, provided detection specificity was not adversely affected.

We began by examining the relationship between measurements of total CA 19-9 and CA 19-9 on individual proteins (Figure 11a). In some cases, patients that were low in total CA 19-9 were distinguishable from pancreatitis patients by their CA19-9 level MUC16. A threshold could be set for this marker that did not result in the detection of additional pancreatitis patients beyond those already detected by total CA 19-9 (keeping the specificity at 75%) but that allowed the detection of five additional cancer patients. We searched the rest of the measurements for a similar relationship. That is, we fixed a detection threshold for total CA 19-9 at 75% specificity (25% false positive rate), set a threshold for each marker that would not increase the false positive rate, and determined if any false negatives by total CA19-9 were picked up by the marker (Figure 11b).

In the UP sample set, with a total of 177 pancreatic cancer and pancreatitis samples, 11 markers picked up at least one of the samples not detected by total CA19-9 (Figure 12). (One of the samples was picked up using detection with the lectin *Bauhinea Purpurea* (BPL), as discussed below.) Some of the measurements were redundant in the samples that were positive, so we simplified the panel to the minimum number of markers necessary to pick up maximum additional samples. A reduced set of four additional markers, CA19-9 detection on MUC1 (#1095), MUC5ac (#1251), and MUC16 (#830), and BPL detection on MUC5ac (#1251), was identified (Figure 11b). Among the 124 cancer samples in this set, 109 were detectable by total

CA19-9 (88% sensitivity at 75% specificity). Using a combination rule in which an elevation (defined by the threshold determined individually for each marker) in any member of the panel indicates a “case,” and a lack of elevation in all markers indicates a “control,” the marker panel picked up eight of the remaining 15 false negative cancer samples (94% sensitivity at 75% specificity) (Figure 11b).

We investigated whether this strategy of combining markers and the resulting biomarker panel were consistent in the repeat dataset of the 177 samples. The correlations between the repeat datasets were good for all the CA 19-9 measurements ($r > 0.8$, Pearson’s r correlation coefficient). When we applied the panel selection strategy described above to the repeat dataset, the same markers were selected, and each additional marker picked up the same false negative samples in both sets (Figure 11c). This result reflects the good reproducibility in the measurements and confirms the accuracy of the panel selection.

We also asked whether the thresholds for each marker determined from one set could be applied to the other set to achieve the same improvement in sensitivity. This analysis partially mimics the implementation of a biomarker panel in clinical settings, in which a pre-determined threshold is used to classify incoming samples. The signal intensities of each marker were median-centered within each dataset to provide a common baseline between the two datasets, and a threshold was determined for each marker in each set using the strategy described above. The thresholds were applied to the opposite set, and the resulting level of discrimination was assessed. The same eight samples were picked up by the marker panel in each set, and seven of those eight were detected by the identical markers regardless of which set the threshold was derived from (Figure 11c). This result supports the reproducibility of the marker panel selected for this sample set and confirms the reliability of the panel selection strategy.

The primary images from selected samples provide further insights into the differences between the samples (Figure 13). The “true positive” sample (from a cancer patient with a total CA 19-9 elevation) shows strong signal at the CA 19-9 antibody and the proteins captured by several other antibodies, as does the “false positive” (from a pancreatitis patient with a total CA 19-9 elevation). The “false negatives” (from cancer patients with no total CA 19-9 elevation) that were picked up by the marker panel show high signals at the respective member of the panel that provided detection. The false negatives that were not picked up by the panel do not show such elevations, indicating that another marker will be required to detect these patients. These results support the concept that by measuring the CA 19-9 antigen on individual proteins, additional patients can be detected that would otherwise be missed by total CA 19-9 measurements. In addition, these data suggest the existence of distinct subgroups of patients defined by the proteins that carry the CA 19-9 antigen.

The set of 177 samples had been run with detection using BPL and Wheat Germ Agglutinin (WGA), and we included those data in the analysis as a preliminary look at the relationships between the CA 19-9 antigen and other glycans. Of note, the false negative that was picked up by BPL detection on MUC5AC (sample LC3607) showed no signal at the CA 19-9 capture antibody when detected with BPL (Figure 13), indicating no overlap between proteins bearing the BPL target and proteins bearing the CA 19-9 antigen. In fact, very little of the CA 19-9 antigen was present in this sample, since detection with CA19-9 showed negligible signals on all spots (Figure 13). This comparison suggests the importance of detecting other glycans besides the CA19-9 antigen for further performance improvement, especially in the cancer patients with no CA19-9 present.

4.3.3 Consistent improvement in sensitivity in independent sample sets

Our next objective was to see if this relationship held up in the other sample sets, which would confirm its generality. The mucins MUC1, MUC5AC, and MUC16 were targeted in the other two datasets, but the same antibodies as in UP sample set were not always used. Therefore we could test the usefulness of detecting CA 19-9 on those proteins independent of any particular antibody. We applied the same strategy as above to determine if any false negatives, as defined by CA 19-9, could be picked up by detecting CA 19-9 on individual proteins, without increasing the false positive rate. The ENH set was run in separate experiments comparing early-stage cancer to pancreatitis and late-stage cancer to pancreatitis (Figure 14a). Of the 17 late-stage cancer patients, all three patients that were low in total CA 19-9 were picked up by the panel (two samples were inconclusive due to missing data), and of the 33 early-stage cancer patients, two of six were picked up. In the UM set, one of the four false negatives out of 40 cancer subjects was picked up (Figure 14b). Between the sets, 25-100% of the false positive samples as defined by total CA 19-9 were detected using CA 19-9 on individual proteins, which is consistent with the result of 7/15 (47%) picked up in the repeat UP sets. Two antibodies contributed to detection in the ENH and UM sets that were not used in the UP set, targeting MUC1 (#1093) and MUC5ac (#831), showing that this result is not dependent on particular antibodies. Therefore in both the ENH and the UM sample sets, sensitivity was improved without reducing specificity using a marker panel of CA 19-9 measurements on individual proteins.

We summarized the performance of discriminating pancreatic cancer from pancreatitis for total CA19-9 and panels of CA 19-9 measurements on individual proteins (Table 3). These results comprise 209 samples from cancer patients (100 early-stage and 109 late-stage) and 112

samples from pancreatitis patients, collected at three different institutions and analyzed in five independent experiment sets. At a fixed specificity of approximately 75%, an improvement in sensitivity (from 2.5% to 15.8%) was observed in all experiment sets. This improvement led to an elevation of the sensitivity of cancer detection from 78-84% by total CA19-9 to 81-100% by the panel.

4.3.4 Size distribution of CA 19-9 carriers in the subgroups

Based on these results, it appears that pancreatic cancer patients consistently fall into the following subgroups: 1) high in total CA 19-9; 2) low in total CA 19-9 but detectable using CA 19-9 on individual carrier proteins; and 3) undetectable by any CA 19-9 measurements. Of fundamental interest is the distribution of CA 19-9 carrier proteins in these subgroups. An approach to visualize the range of proteins carrying the CA 19-9 antigen is to fractionate the serum proteins using SDS-PAGE and immunoblot for the CA 19-9 antigen, which we did for representative samples from the subgroups defined above (Figure 15). The samples that were high in CA 19-9 by microarray showed a broad range of molecular weights with high signal, indicating many proteins containing the CA 19-9 antigen. The samples that were false negative by CA 19-9 but detected by the panel showed only faint bands at high molecular weights (>150 kD) and the samples not detected by any marker showed no discernable or only faint bands. This results shows that no major protein carriers of the CA 19-9 antigen, at least in the molecular weights observed in this format, are present in the low CA 19-9 samples. Thus, the detection of the remaining samples not picked up by the panel most likely will rely on additional proteins or glycans.

4.4 Discussion

The need for improved blood markers for pancreatic cancer is great. Such markers would have important applications in the detection and diagnosis of the disease, leading to improved patient management and outcomes. The sub-optimal performance of the CA 19-9 assay may, in some cases, be due to the appearance of the CA 19-9 carbohydrate antigen on carrier proteins that are not specific to cancer. By detecting the antigen specifically on the proteins that are the predominant carriers in cancer, improved performance may result. We examined this possibility using antibody arrays with glycan detection, which provided a convenient approach to measuring the CA 19-9 antigen on multiple, individual proteins. We found that the mucins MUC1, MUC5AC, and MUC16 are indeed major cancer-associated carriers of CA 19-9, based on the strong discrimination of cancer from pancreatitis using those individual measurements. Moreover, the individual measurements are complementary to total CA 19-9, since some patients that are low in total CA 19-9 can be detected by the individual measurements, using detection thresholds that do not increase overall detection specificity. The end result is an improvement in the sensitivity of cancer detection. The fact that the same result was achieved in three sample sets from three different institutions, and with more than one antibody per protein, supports the generality of this finding.

The ability to more sensitively detect cancer relative to benign disease conditions could enable more rapid and accurate diagnoses of patients with suspected cancer. A possible area of application would be among patients that have pancreatic abnormalities as discovered by CT scan. A wide variety of conditions in addition to malignancies produce abnormal pancreatic findings by CT [121], such as cystic lesions, pancreatitis, and common bile duct stones, and only some require further intervention. Because no molecular marker exists to sort out the conditions,

nearly all patients go on to endoscopic ultrasound and potential biopsy. A reduction in this invasive, costly, and risky procedure is desirable, considering the high rate of patients with benign conditions that receive it. A molecular marker that sensitively detects cancer, so that a negative result safely rules out patients that do not need further workup, could address this need. The gain in detection sensitivity made here is a step in that direction. Future work includes further validating and characterizing the improved sensitivity of our new marker panel and determining the panel's ability to meet the performance needs of specific clinical applications. It may be the case the certain clinical applications will have a requirement for greater specificity, in which case the threshold could be elevated to reduce the false positive rate. The improvements observed using CA 19-9 on individual proteins should hold at other thresholds.

The finding of apparent subgroups of patients based on CA 19-9 protein carriers has implications for the development of a biomarker panel. Clearly some of the patients express CA 19-9 many proteins, some on only a small number of proteins, and others not at all. For those patients expressing the antigen on few proteins, the detection of CA 19-9 on the proper carrier allows detection probably because of a concentration of the signal at that protein. The Western blot results (Figure 15) showed that not many proteins in these samples carried the antigen, explaining the low total CA 19-9 signal and the need for detection on a specific protein in order to pick up the signal. Furthermore, since the mucin proteins seem to be highly cancer-associated carriers of the antigen, very few pancreatitis patients showed elevations of CA 19-9 on the mucins (Figure 10, Figure 11b), which allowed the use of lower thresholds to more sensitively pick up the cancer patients.

For those patients not showing elevations in CA 19-9 on any of these mucins, we will need either additional proteins or additional glycans. The existence of other protein carriers of

the CA 19-9 antigen was suggested by the fact that some of the “false negative” patients showed discernable CA 19-9 signals on the microarrays and Western blots. Therefore, other carrier proteins are present, which if targeted might enable specific detection of these patients. The challenge is to identify those carrier proteins. Other patients may indeed secrete mucins into the blood, but mucins that do not carry the CA 19-9 antigen. This situation was represented by patient 3607 (Figure 11, Figure 13). This patient showed no CA 19-9 signal on any carrier, but strong signal at the MUC5AC capture antibody when detected by the lectin BPL. The glycan bound by BPL, terminal galactose, is distinct from the glycan bound by CA 19-9, confirming the need for the detection of additional glycans beyond CA 19-9. This result is consistent with the fact that certain individuals, estimated around 5% of the population, are genetically deficient in an enzyme that completes a critical step in the biosynthesis of the CA 19-9 antigen. Detection with BPL may provide the ability to detect some of this population.

Another approach to enhancing the ability to detect the cancer patients is to improve the limit of detection of the analytical assay. Some of the patients not detected in this study may have mucin proteins secreted into the blood but at very low levels, which might be detectable given a very sensitive assay. Several options are available for improving the detection limits of the assay. Amplification of the fluorescence signal is possible using rolling-circle amplification [122, 123] or tyramide signal amplification [124]. A novel format that restricts the sample to ultra-low detection volumes can lower detection limits using enzyme-based chemiluminescence detection [125]. A new generation of electrochemical biosensors is achieving or surpassing detection limits achieved by fluorescence [126], which provides another possible route for the improved detection of cancer patients.

The approach to biomarker development demonstrated here has two significant aspects that may be useful in a range of projects. One is the detection of glycans on specific proteins. Greater accuracy for particular disease states can be achieved measuring glycans on specific proteins, rather than just protein levels, as with standard immunoassays, or just the levels of a particular glycan on all proteins, as with the conventional CA 19-9 assay. The antibody-lectin sandwich array provides an ideal format for testing combinations of proteins and glycans for biomarker performance [127]. The proteins and glycans to be targeted on the arrays can be derived from known molecular alterations, such as mucins in pancreatic cancer [85, 128, 129], or from genomics, proteomics, and glycoproteomics studies. Glycoproteomics methods used in combination with antibody arrays could represent a powerful strategy for biomarker development [130], the former providing potential new proteins and glycans to test, and the latter providing an efficient and accurate means of testing multiple candidates.

The other significant aspect of the approach demonstrated here is the targeting of predefined subgroups for the identification of complementary markers. Studies of cancers of other organs have identified subcategories of disease defined by molecular characteristics [131]. Although gene expression and other types of molecular profiling studies have been performed on pancreatic ductal adenocarcinoma [132, 133], clear subcategories of the disease have not emerged. However, it is likely that defined subgroups of the disease exist that have distinct molecular characteristics and that produce distinct alterations in the blood. Therefore, we would expect that a panel of complementary markers, each member specific for a subgroup of disease, would be required to detect a high percentage of patients. Discovery efforts focused on such subpopulations could be more effective than efforts aimed at the whole population, since markers specific for a small subgroup would be diluted over the population. Furthermore, simple

combination rules to bring together the complementary markers may faithfully represent the actual biological entities, leading to accurate overall classifications. Future work to further improve the biomarker panel found here will focus on finding additional complementary markers specifically in the subgroup of patients remains misclassified.

This work gives encouragement to the possibility of accurate, early diagnosis of pancreatic cancer, using a marker panel that detects nearly all cancer patients with a low rate of false positive detection of pancreatitis. The improvement over the conventional CA 19-9 assay was achieved by detecting the CA 19-9 antigen on specific mucin proteins rather than on all protein carriers. The sensitivity of cancer detection was raised to 84-100% from 79-84%, at 75% specificity, in three independent sample sets from three different institutions. The clinical implementation of this marker could involve assisting doctors in the differential diagnosis of benign and malignant disease or determining the course of additional diagnostic workup. Further validation will be performed using blinded samples collected from the setting of the intended clinical application, in accordance with the developed standards for biomarker validation [134].

4.5 Materials and Methods

Serum and plasma samples. Serum samples from pancreatic cancer, pancreatitis and healthy subjects were collected at Evanston Northwestern Healthcare and the University of Michigan Medical School, and plasma samples were collected at the University of Pittsburgh School of Medicine (Table 1). All sample collection was under protocols approved by a local Institutional Review Board. The control subjects had no pancreatic lesions. All samples were stored at -80°C and sent frozen on dry ice. Each aliquot had been thawed no more than three times before use.

Antibodies and lectins. The antibodies and lectins were obtained from various sources (see Error! Reference source not found.). All antibodies were screened for reactivity and integrity using Western blots, purified, and prepared at 0.5 mg/ml in pH 7.2 phosphate-buffered saline (PBS) for non-contact array printer and at 0.25 mg/ml in pH 7.2 PBS for contact array printer. The steps of antibody purification included ultracentrifugation at 47,000g at 4 degree for 1 hour and dialysis (Slide-A-Lyzer Mini Dialysis Units, Pierce Biotechnology) against pH 7.2 PBS at 4 degree for 2 hours.

Microarray fabrication. Approximate 170 pg (350 pl at 500 µg/ml or 700 pl at 250 µg/ml) of each antibody was spotted on the surfaces of ultra-thin nitrocellulose-coated microscope slides (PATH slides, GenTel Biosciences) by a piezoelectric non-contact printer (Biochip Arrayer, PerkinElmer Life Sciences) for the slides used in ENH and UM sets, and by a non-contact microarrayer (sciFLEXARRAYER, Scienion) performed at GenTel Biosciences (Madison, WI) for the slides used in UP set. Forty-eight identical arrays containing triplicates of all selected antibodies were printed on each slide. The three spots in a triplicate were positioned adjacent to each other in the slides used in ENH and UM sets, and the position of each spot in triplicates was randomly assigned in the slides used in UP set. Hydrophobic borders were imprinted around each array using a stamping device (SlideImprinter, The Gel Company, San Francisco, CA).

Microarray assays. Microarray sandwich assays were performed to measure either the level of total CA19-9 or the glycan levels on the selected proteins captured by the immobilized antibodies (Figure 9a). The sandwich assay consisted of four 1-hour-incubations in room temperature (RT) with the following reagents: 1) blocking buffer (PBS containing 0.5% Tween-20 (PBST0.5) and 1% BSA); 2) a serum sample, diluted two-fold in 1XTBS containing 0.08%

Brij, 0.08 Tween-20, 50 µg/ml protease inhibitor cocktail (Complete Protease Inhibitor Tablet, Roche Applied Science), and a cocktail of IgG from mouse, goat, and sheep each at 100 µg/ml and rabbit IgG at 200 µg/ml (Jackson ImmunoResearch Laboratories, Inc.); 3) biotinylated detection antibody or lectin (2 µg/ml), diluted in PBST0.1 containing 0.1% BSA; 4) streptavidin-phycoerythrin (10 µg/ml, Roche Applied Science), diluted in PBST0.1 containing 0.1% BSA. After each step, the slides were rinsed in three baths of PBST0.1 and dried by centrifugation (Eppendorf 5810R, rotor A-4-62, 1500 x g).

The measurement of glycans by using lectins detection on the captured proteins (Figure 9a) was carried out as above, except the glycans on the spotted antibodies were derivatized to prevent lectin binding to the antibodies [74], and the arrays were probed with glycan-binding lectins.

Fluorescence emission from the phycoerythrin was detected at 570 nm using a microarray scanner (LS Reloaded, Tecan). All arrays within one slide were scanned at a single laser power and detector gain setting. The images were quantified using the software program GenePix Pro 5.0 (Molecular Devices, Sunnyvale, CA). Spots were identified using automated spot-finding or manual adjustments for occasional irregularities. The local backgrounds were subtracted from the median intensity of each spot, and triplicate spots were averaged using the geometric mean. The coefficient of variation between replicate analyzed spots was 10-15%.

Statistical analyses and software. Pearson correlations, Student's T-tests, and receiver-operator characteristic analyses were calculated using Microsoft Excel. The scatter and box plots were created using OriginPro 8, and figure production was performed using Canvas 8. Clustering and visualization were performed using the programs Cluster and Treeview and MultiExperiment Viewer.

APPENDICES

TABLES

Set #	Set provider	Early-stage cancer **	Late-stage cancer ***	Pancreatitis	Healthy	Total
1	Evanston Northwestern Healthcare (ENH)	33	20	26 / 39	20	
2	University of Michigan (UM)		40	24	24	420
3	University of Pittsburgh (UP)*	58	58	49	43	

* This set of 220 samples were double-blinded at the time of experiment

** Stage I, II

*** Stage III, IV.

Table 1 Sets of plasma sets used in the study.

Marker	Antibody	ENH2-early	ENH2-late	UM2	UP	UP repeat	Average
Total CA19-9	341	0.81	0.81	0.91	0.87	0.88	0.86
	1093	0.82	0.93	0.89	0.81	0.8	0.85
CA19-9 on MUC1	340	0.65	0.86		0.71	0.66	0.72
	684	0.52	0.8				0.66
	1094	0.52	0.84				0.68
	1095				0.68	0.63	0.66
	831	0.77	0.92	0.75	0.81	0.8	0.81
CA19-9 on MUC5ac	1091	0.8	0.86	0.91	0.63	0.52	0.74
	1092	0.55	0.87				0.71
	1251				0.79	0.78	0.79
	339	0.58	0.8				0.69
CA19-9 on MUC16	830	0.62	0.82	0.9	0.86	0.85	0.81
	1098	0.66	0.8	0.9	0.82	0.78	0.79
	1099	0.51	0.81	0.83	0.81	0.78	0.75

* The area-under-the-curve is given

Table 2 Performance of individual markers for cancer from pancreatitis in each sample set.

Compared to all BENIGN	UP ₁ ⁻ #1*	UP ₁ ⁻ #1**	UP ₂ ⁻ #2*	UP ₂ ⁻ #2**	UP ₂ ⁻ #2**	ENH _{early} ⁻ early*	ENH _{early} ⁻ early**	ENH _{late} ⁻ late*	ENH _{late} ⁻ late**	UM*	UM**
False positives	12	12	12	12	12	6	6	9	9	6	6
All true benign	48	48	49	49	49	26	26	39	39	24	24
Specificity	75%	75%	75.5%	75.5%	75.5%	76.9%	76.9%	76.9%	76.9%	75%	75%
False negatives	15	8	19	12	12	7	5	3	0	4	3
All true cancer	124	124	128	128	128	33	33	19	19	40	40
Sensitivity	87.9%	93.5%	84.4%	90.6%	90.6%	78.8%	84.8%	84.2%	100%	90%	92.5%
Performance (accuracy)	84.3%	88.3%	81.9%	86.4%	86.4%	78%	81.4%	79.3%	84.5%	84.4%	86%
% of FN picked up	47%(7 of 15)	36.8%(7 of 19)	28.6%(2 of 7)	100%(3 of 3)	25%(1 of 4)						

* Measured by total CA19-9 only **Measured by panel

Table 3 Comparison of the performance of total CA 19-9 and the marker panel in each set.

FIGURES

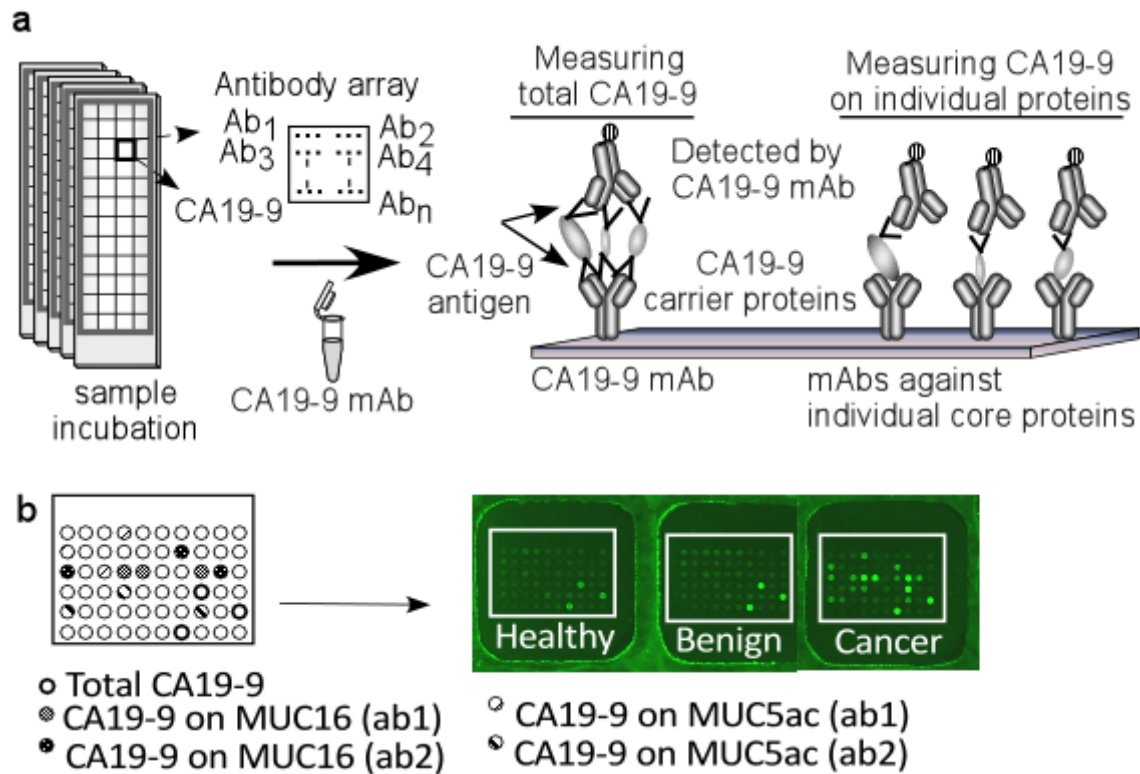


Figure 9 Detection of total CA19-9 and CA 19-9 on individual proteins using antibody arrays.

a) High-throughput sample processing and array-based sandwich assays for CA19-9 detection. Forty-eight identical arrays are printed on one microscopic slide, segregated by hydrophobic wax boundaries (left). A set of serum samples are incubated on the arrays in random order, and the arrays for the entire sample set are probed with the CA 19-9 detection antibody. Total CA19-9 is measured at the CA19-9 capture antibody, and CA19-9 on specific proteins is measured at the the individual antibodies against those proteins (right). b) Representative raw image data from each of the sample groups. Triplicates of each antibody were randomly positioned on the array.

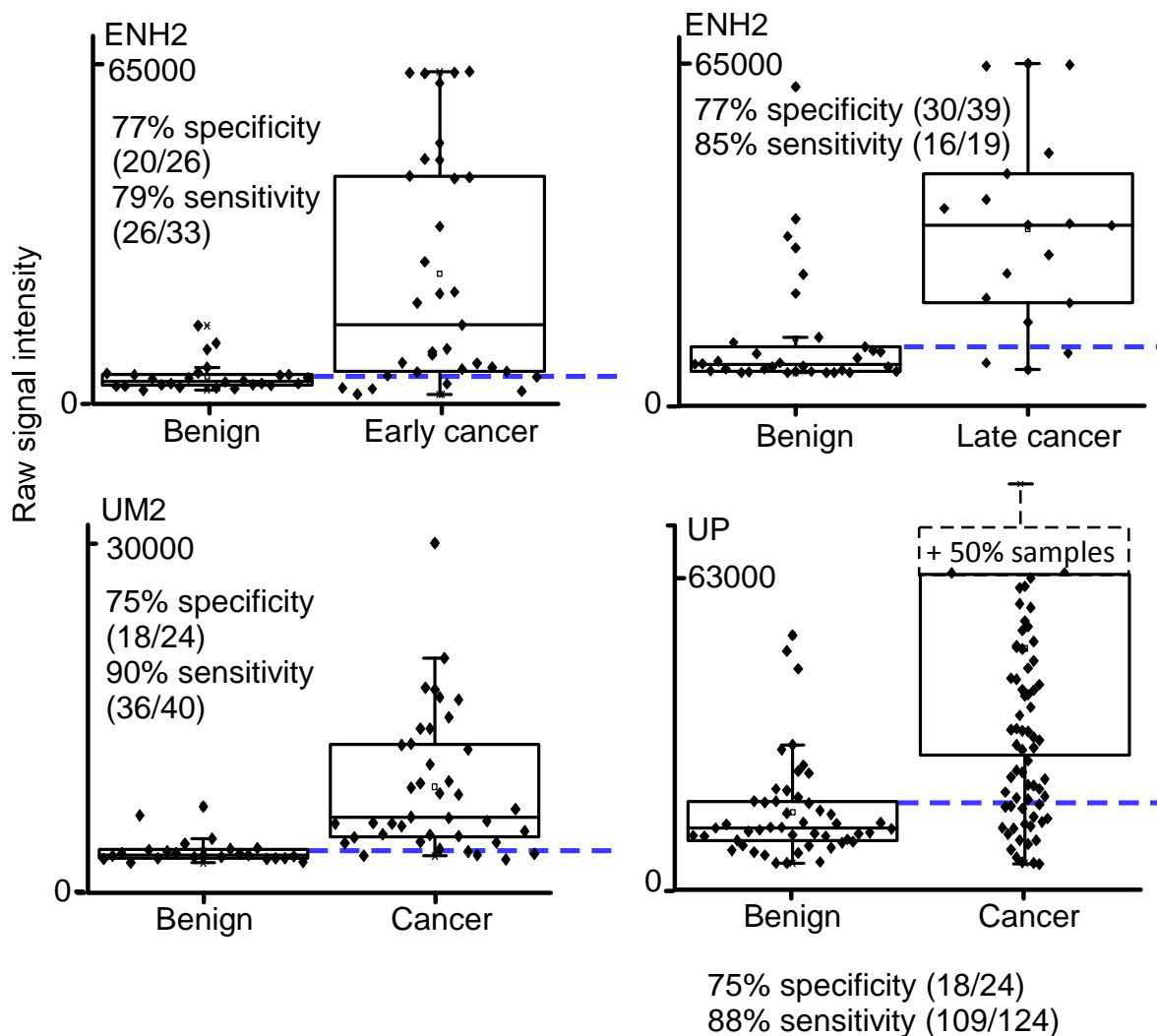


Figure 10 Distribution of total CA19-9 levels in pancreatic cancer and pancreatitis patients.

Each point represents an individual sample. The boxes indicate the quartiles, with the median indicated by the solid horizontal lines, and the vertical lines mark the ranges. Blue dash lines indicate the threshold selected for further analysis.

a) Comparison of CA19-9 on MUC16 to total CA19-9. The levels of CA 19-9 on MUC16 for each sample are plotted along the vertical axis, and the total CA 19-9 levels for the same samples are plotted along the horizontal axis. The plot shows only the lower 50% of the samples by total CA 19-9. The vertical line indicates the threshold defined to give 75% specificity by total CA19-9. The horizontal dashed line indicates a proposed threshold for CA19-9 on MUC16 which would result in the detection of additional cancer samples (noted by the arrows) without detecting additional pancreatitis samples. b) Combined results of total CA19-9 and four additional complementary markers. The samples are ordered in the columns (Bn is benign, EarlyC is early-stage cancer, LateC is late-stage cancer, Cancer is unknown stage cancer) and the markers in the rows. The threshold for total CA19-9 was set to 75% specificity, and the threshold for each additional marker was defined as in panel a. A yellow square indicates a measurement above the threshold, a black square indicates below the threshold, and gray squares are missing data. The blue box denotes the cancer samples not detected by CA 19-9 (CA 19-9 measurements in the red box). The samples picked up by the additional markers are highlighted by blue column labels. c) Comparisons of panel performance in duplicate sets. The panel developed from each experiment was applied to both the original and the duplicated set. The marker that detected each sample (indicated in the rows) is given for each application of the marker panels. The markers that were elevated were highly consistent.

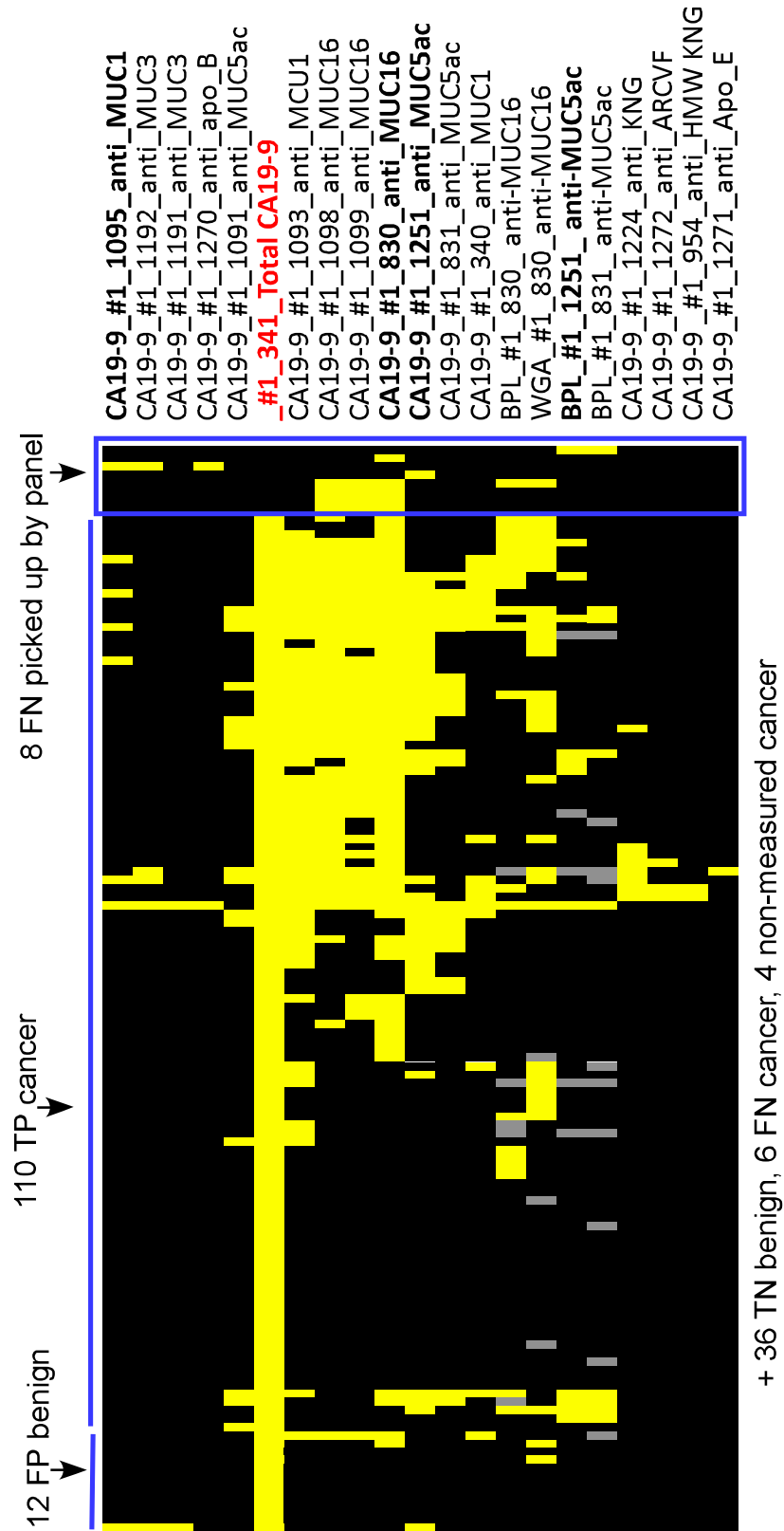


Figure 12 Combined results for all markers tested in UP set.

The samples are ordered in the column and markers in the rows. Marker name is labeled as “detection_experiment set (#1)_capture antibody ID_anti_capture antibody target. Markers selected in the panel is highlighted in bold. Total CA19-9 is in red. TN, true negative; FN, false negative; TP, true positive; FP, false positive. All defined by total CA19-9. Yellow / black square, measurement above / below the threshold.

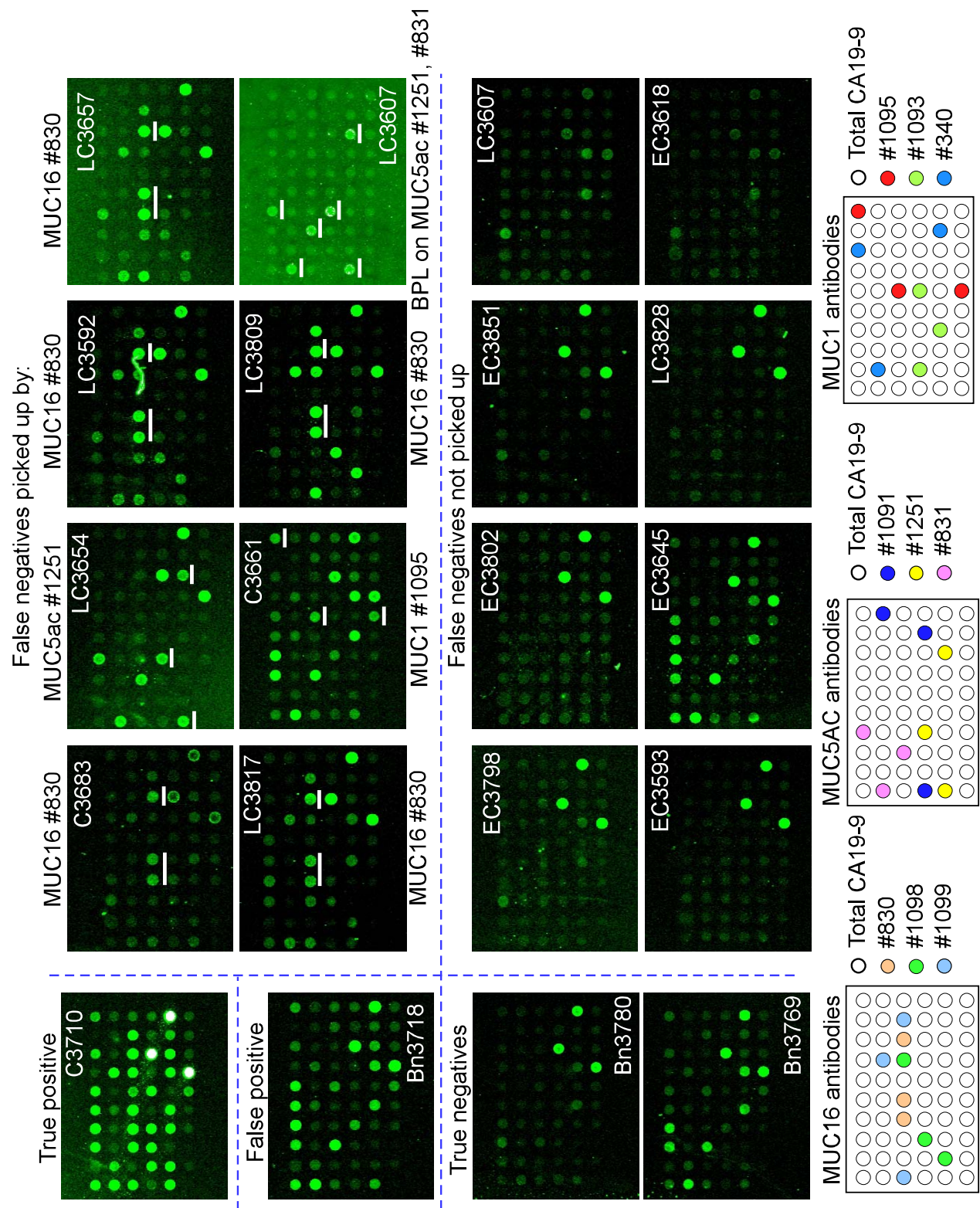


Figure 13 Raw images of arrays from subgroups defined by total CA19-9.

Cancer samples that were detected by CA 19-9 (true positive), not detected by CA 19-9 (false negative) but picked up by the panel, or not detected by CA 19-9 or the panel are represented. In addition, pancreatitis samples that were not detected by CA 19-9 (true negative) or detected by CA 19-9 (false positive) are represented. The sample identifier is given within each array. In the subgroup picked up by the panel (top-right), the antibody used to detect a given sample is listed adjacent to each array. The corresponding antibody spots are underlined white. Two arrays for sample LC3607 are shown, one detected with BPL (rightmost column, row 2), and the other detected with CA19-9 (rightmost column, row 3). All other arrays were detected with CA19-9. The bottom panels show maps of antibodies targeting MUC16 (left), MUC5AC (middle) and MUC1 (right).

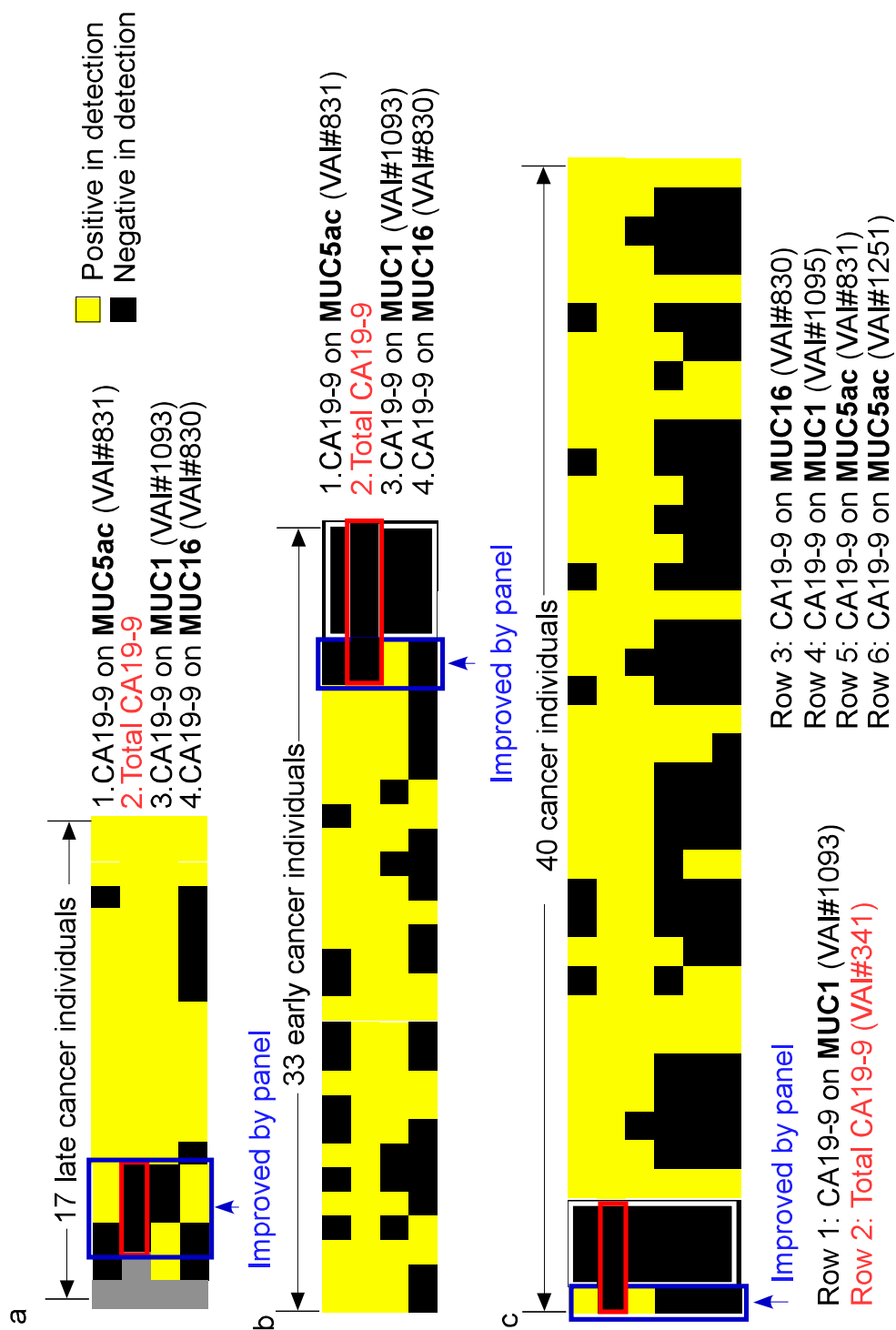


Figure 14 Panel performance in additional sample sets.

Marker selection was performed for sample sets 2 and 3 as described above. Specificity was fixed at approximately 75% by total CA19-9, and for each additional marker, the threshold

was defined as in Figure 11a. The yellow squares indicate measurements above the threshold for a given marker, black indicates below the threshold, and gray indicates missing data. Each column represents an individual sample from the relative set, and the row indicates the marker used with antibody ID followed in parenthesis. The blue boxes highlight the sample(s) picked up by the panel, and the red and white boxes indicate the false negatives defined by total CA19-9 and the panel, respectively. a-b) Sample set 1, comprising late-stage (a) and early-stage (b) cancer patients. c) Sample set 2, comprising a mix of early and late stage patients. In both sets, the samples from the pancreatitis patients are now shown.

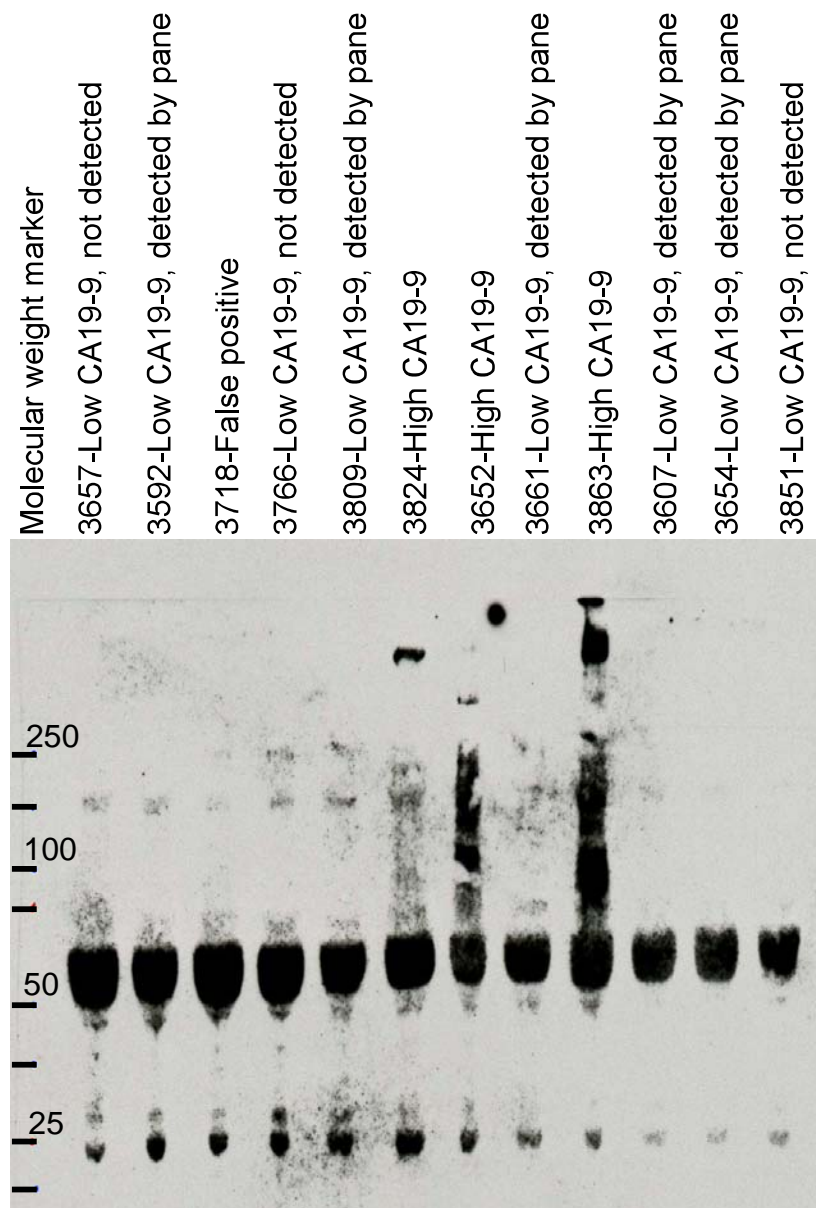


Figure 15 CA19-9 western blot of individual samples.

The number on top indicates sample ID corresponding to Figure 13.

CHAPTER 5 NEW CA19-9 CARRIER IDENTIFIED IN PANCREATIC CANCER

T. Yue, P.Andrews, and B.B. Haab. Identification of novel CA19-9 carrier in pancreatic cancer.
(In preparation for submission to BMC Cancer)

5.1 Abstract

The current standard serum marker for pancreatic cancer, CA19-9, shows elevations in some patients with chronic pancreatitis condition, making it unreliable for the diagnosis of pancreatic cancer. Since the assay of CA19-9 detects the carbohydrate antigen on multiple protein carriers, we hypothesized that the detected CA19-9 comprising of signals from both the cancer-associated carriers and the cancer-unrelated carries. Therefore, by identifying cancer-associated carriers and detecting the antigen on these carriers, we may enhance the cancer detection. We immunoprecipitated the CA19-9 carriers from separated high-CA19-9 sample pools selected from pancreatic cancer and non-cancer groups. We used the LC MALDI-TOF mass spectrometry to identify potential protein carriers from the sample pools, and validated the candidates using antibody sandwich assays. Apolipoprotein B-100 (ApoB) was found as CA19-9 carriers in around 25% of the pancreatic cancer patients, however, the CA19-9 on ApoB does not contribute to the CA19-9 elevation associated with pancreatic cancer. ApoB represents a group of CA19-9 carriers that are cancer-unrelated. This explains why the CA19-9 detection can be improved by incorporating CA19-9 detection on selected carriers, and strengths the importance of future work in identifying such carriers.

5.2.Introduction

Improved diagnostic methods for pancreatic cancer are greatly needed. Pancreatic cancer often advances to an incurable stage prior to detection, leading to very short survival time for patients. Furthermore, difficulties in distinguishing benign from malignant disease and predicting optimal treatment courses can lead to sub-optimal management of patients. Molecular diagnostics methods that can provide accurate detection of early-stage cancer or information

about disease extent could lead to more effective treatment of patients and overall better outcomes. Thus far, serological tests that meet this need have been elusive.

The current best serological marker for pancreatic cancer is the CA 19-9 assay. The use of CA 19-9 for a wide variety of purposes has been extensively investigated, including early detection, diagnosis, prognostication, and monitoring of tumor responses and recurrence [116, 135-137]. It is elevated in the blood of about 80% of pancreatic cancer patients, and its levels often follow the regression or progression of tumors in patients receiving therapy for pancreatic cancer. Accordingly, the marker is commonly used to follow patients receiving treatment for pancreatic cancer. Blood levels of CA 19-9 also can be elevated in conditions including liver damage, bile duct obstruction, and pancreatitis. Because of those additional causes of elevation, the CA 19-9 test is not specific enough to be used for pancreatic cancer detection or diagnosis. In addition, the sensitivity for cancer is not sufficiently high to rule out cancer upon a low reading. Improvements to the CA 19-9 assay are clearly needed for a better control of pancreatic cancer.

The CA 19-9 assay measures a carbohydrate antigen that is found on many different proteins. The repertoire of these “carrier” proteins is not well defined but is known to include mucins and other adhesion molecules such as carcinoembryonic antigen [119]. It also is not known whether the composition of the CA 19-9 carrier proteins is different between disease states, for example whether the proteins bearing the CA 19-9 antigen in pancreatitis are different from those bearing the CA 19-9 antigen in pancreatic cancer. If the carrier proteins are different, an improved discrimination between the disease states may be possible by measuring the CA 19-9 antigen on specific proteins, rather than on all proteins [85]. The antibody-lectin sandwich array platform developed earlier is ideal for testing that concept, since the levels of a particular glycan epitope can be measured on many different proteins in parallel (Fig. 1a) and compared

across many different samples. In order to implement such experiments, one must first define the potential carrier proteins to target on the antibody array.

Therefore the goals of this study were to identify potential blood-based carrier proteins of the CA 19-9 antigen and to investigate whether the composition of the protein carriers is different between pancreatic cancer and pancreatitis. We first examined whether the rate at which the CA 19-9 antigen is elevated on known mucin carriers is different between the two patient groups. This analysis gave insight into the concept that the protein carriers of the CA 19-9 antigen are distinct between disease states and provided the motivation for efforts to discover additional carriers of the CA 19-9 antigen. Next, to identify such carrier proteins, we immunoprecipitated the CA 19-9 antigen from the blood of patients from each of the disease groups and identified the captured proteins using mass spectrometry. Finally, we finally confirmed that selected proteins bear the CA 19-9 antigen, and we examined the prevalence of the antigen carried by these candidates in the patient groups.

5.3 Results

5.3.1 Variation in the carriers of the CA 19-9 antigen in patient subpopulations

We began by examining known carriers of the CA 19-9 antigen, which included the mucin proteins MUC1, MUC5AC, and MUC16 [119], to investigate whether the distribution of the CA 19-9 antigen among these proteins vary among patient groups. We used antibody arrays to explore this question, by which the CA19-9 glycan could be measured for its total level as well as its level on individual proteins simultaneously (Figure 16a). For the measurement of total CA19-9, a monoclonal antibody against CA19-9 was used as both the capture and detection

antibody. To measure CA19-9 on specific proteins, antibodies targeting the potential carrier proteins were immobilized on the microscopic slide, and the CA 19-9 antibody was used for detection. In a standard antibody array platform, which can easily incorporate dozens of capture antibodies within one array [138], both the total amount of CA 19-9 and the specific levels of CA19-9 on particular proteins can be measured all in one assay.

We took measurements on samples from three disease groups, including pancreatitis, early-stage pancreatic cancer and late-stage pancreatic cancer patients (Figure 16b-d). To look at the prevalence of the CA 19-9 antigen on each of the mucins in each group, we set a threshold for each antibody to determine which of the individual measurements were elevated, relative to a control group. Good general agreement was observed between antibodies targeting the same protein, which supports the specificity of the detection. (The specificity and reproducibility of these assays also had been confirmed in previous work [119, 139].) In the pancreatitis group (Figure 16b), about 25% of the patients showed elevations in CA 19-9 on all three mucins, and slightly less than half showed no or sporadic elevations. About a third of the patients showed elevations only on MUC16. In early-stage cancer, again about 25% of the patients have CA 19-9 elevations on all mucin proteins and another 25% have no or few elevations. The rest of the patients show different patterns than the pancreatitis group. A subset of about a third of the patients showed elevations only on two of the MUC5AC antibodies and one of the MUC1 antibodies, and another subset of about 15% (5 of 33) of the patients have elevations on those antibodies plus half of the MUC16 antibodies. The distinctions between antibodies targeting the same proteins suggest the existence of discrete isoforms of these mucins that are preferentially expressed in early-stage cancer. Late-stage cancer showed yet another pattern. Over half (11 of 19) showed elevations on all mucins, and another ~25% (5 patients) showed variable elevations

on each of the mucins. A distinct subset of three patients showed elevations only on MUC1. These results support the concept that subgroups of patients have distinct carrier proteins of the CA 19-9 antigen, and that the carrier proteins may be different between disease states.

5.3.2 Identification of protein carriers of the CA 19-9 antigen

Based on the above result, we sought to identify additional potential carriers of the CA 19-9 antigen in pancreatic cancer and pancreatitis patients. We used a glycoproteomics strategy based on immunoprecipitation and mass spectrometry (Figure 17). For each of four groups, healthy controls, pancreatitis patients, early-stage pancreatic cancer patients, and late-stage pancreatic cancer patients, a sample pool was created from five to ten samples each containing a high level of CA19-9. Next, the CA19-9 carrier proteins were isolated from each pool by immunoprecipitation with the CA 19-9 monoclonal antibody, followed by deglycosylation using a combination of glycosidases that cleaved both N- and O-type of glycans from the pulled-down glycoproteins. Deglycosylation enabled us to obtain more protein hits identified in the mass spectrometry (data not shown). In the following steps, the pools of CA19-9 carrier proteins obtained from each sample group were further separated by one-dimensional sodium dodecyl sulfate gel electrophoresis (1D-SDS PAGE) for the identification of proteins by tandem mass spectrometry (MS), using matrix-assisted laser desorption/ionization with time-of-flight detection (MALDI/TOF/TOF). For selected identifications from the MS analysis, antibodies were obtained and used to measure the CA 19-9 levels on the target proteins in independent samples.

Three independent experiment sets for MS analysis were assembled. From 275 to 500 high-confidence protein identifications were achieved for the sample pools. Among these, 60 hits representing 21 distinct proteins were identified by more than one peptide and showed some

differences of the number of peptides between pancreatic cancer and non-cancer pools in at least one experiment. Four proteins targets were particularly noticed (Table 4, Table 5).

Apolipoprotein B (ApoB) was one of the most abundant and consistent hits identified in all three experiments. In two of the three experiments, ApoB was found with more hits identified in cancer than in pancreatitis or healthy. Another protein from the apolipoprotein family, ApoE, was also noticed. This protein closely associated with ApoB for function [140], it remained undetectable in benign in all three experiments, but was found in cancer for once. Armadillo Repeat gene deleted in Velo-Cardio-Facial syndrome (ARVCF) was noticed for the reason that none of its peptides were detected in either benign or healthy from all three experiments, but they were found in both early- and late-stage cancer in one of the experiments. For Kininogen, similar to ARVCF and ApoE, though it was found in only one experiment, higher numbers of peptides hits were identified in cancer for both high- and low-molecular-weight isoforms of it. Monoclonal antibodies targeting the above candidates were obtained and immobilized on microscopic slides for further validation using microarray.

5.3.3 Prevalence of the CA 19-9 antigen on ApoB in each disease state

We used the antibody array method depicted above (Figure 16a) to verify that the CA 19-9 antigen is found on the selected proteins and to examine the frequency and levels of that expression. For each of the four selected proteins, namely, ApoB, ApoE, ARVCF and Kininogen, one antibody was used as capture antibody in microarray sandwich assays, and the CA 19-9 levels on the captured proteins was detected among 172 samples (48 benign and 124 cancer). ApoB showed strong signals in selected patients, while the other three proteins showed weaker elevations among fewer patients (Figure 18).

In order to further validate the MS identification of ApoB as a CA19-9 carrier, we first examined the reproducibility of microarray signals by repeating the antibody microarray sandwich assay on the same sample set (Figure 19a). These two measurements were well correlated ($r=0.9$, Pearson's r correlation coefficient). In antibody microarray sandwich assays, ApoB was captured first and CA19-9 level on ApoB was measured next. To validate the microarray signal, we also measured the level in a reverse manner. One sample with high CA19-9 level on ApoB and two samples with low signals were selected based on the results obtained from microarray assays. CA19-9 from each sample was pulled down by Immunoprecipitation using a CA19-9 antibody, followed by the detection using ApoB antibody. Same pattern of signals was observed between the two measurements in which the targets for capture and detection were reversed (Figure 19b). These results confirm the identification of ApoB as a carrier of the CA19-9 antigen in certain samples.

We next examined more closely the frequency with which CA 19-9 is found on ApoB. We first examined the distribution of signals from both sample groups. There was only slightly significant increase of CA19-9 on ApoB in cancer group than that from pancreatitis group ($p=0.044$, Mann-Whitney test) (Figure 20a). We examined the prevalence of ApoB as a CA19-9 carrier in each group to understand its contribution to CA19-9 elevation in cancer. Threshold for CA19-9 was set at 75% specificity, threshold for ApoB was determined from Figure 18. The results indicated that ApoB is a carrier of CA19-9 in only a small portion of subjects in all of the three groups (Figure 20b). In order to understand if ApoB contributed to the elevation of CA19-9 in cancer, we compared the level of CA19-9 on ApoB with that of total CA19-9. Within each group, change of CA19-9 on ApoB had no correlation with the change of total CA19-9 ($r=0.17$ (benign), 0.07 (cancer), Pearson's r correlation coefficient), suggesting that ApoB was not a

major carrier for the patients with high level of total CA19-9 in each group (Figure 20b).

Furthermore, when comparing cancer patients with pancreatitis samples, ApoB was not the major contributor to the elevation of CA19-9 in pancreatic cancer ($R^2=0.03$, coefficient of determination). These results indicated that CA19-9 on ApoB was not a good biomarker for detecting pancreatic cancer from pancreatitis if used alone.

The most important potential application of identifying additional CA19-9 carrier proteins was to help improving the performance of total CA19-9 by detecting false negative cancer samples. The final question we wanted to ask was whether the CA19-9 level on ApoB can still contribute to the standard marker total CA19-9 as in a biomarker panel. Although it is not a good biomarker as individual, in some cases, if a marker was able to pick up any cancer samples that were missed by total CA19-9, it can still contribute to the improvement of detection, for example, the change of the level of CA19-9 on a mucin protein MUC1 captured by antibody 695 clone. This marker did not perform well as an individual marker (Area under the curve (AUC) =0.68, Receiver operator characteristic). However, at a fixed specificity of 75% for total CA19-9, this marker was picking up one false negative cancer out of 15, improved the performance of total CA19-9 by increasing sensitivity without reduction of specificity (data in preparation for submission). However, in the case of measuring CA19-9 on ApoB, there was no improvement (Figure 20c). These results suggest that although ApoB was a confirmed carrier of CA19-9, the change of CA19-9 on ApoB, whether used as individual marker or in complementary to total CA19-9, is not a discriminative indicator of pancreatic cancer from pancreatitis.

5.4 Discussion

The distribution of CA19-9 on multiple proteins is diverse. Profiling of CA19-9 on three mucins (1,5ac,16) from pancreatic cancer and non-malignant pancreatitis revealed distinct pattern of the CA19-9 distribution, which suggests the existence of protein preference in the addition of CA19-9 for a particular disease state. Indeed, even for the same glycoprotein, only certain isoforms of this protein are susceptible for the CA19-9 glycosylation. For example, a previous study has showed that CA19-9 epitope was detected in the recombinant MUC1 protein with defecting tandem repeat (TR) region but not in the MUC1 isoform with 10 TR regions[141]. Similarly, in our study when multiple antibodies were used to target the same mucin, only isoforms captured by certain antibodies bear high level of CA19-9.

Apolipoprotein B (ApoB) was identified as a novel CA19-9 carrier from this study. This lipid-metabolic protein has been the prime target for studying lipoproteins and arteriosclerosis since late-1980s [142] for its central roles in lipoprotein metabolism. Two isoforms of ApoB, ApoB-48 and ApoB-100 were identified [143] and found with separated roles. ApoB-48 is responsible for intestinal absorption of dietary fibers while ApoB-100 is virtually the only protein component of low-density-lipoproteins (LDL). The antibody used in our study to capture ApoB reacted with both isoforms. However, in the detection from CA19-9-enriched protein pools, we found tons of the high-molecular weight ApoB-100 (516kDa) but none of the other (241kDa). This agreed with the mass spectrometry identification, in which multiple peptide hits were repeatedly detected for ApoB-100 but none for ApoB-48, even though both isoforms were secreted to serum. CA19-9 on ApoB-100 presents in around 15%-25% of individuals in all three groups including healthy, pancreatitis and cancer individuals, suggesting that ApoB is unlikely a CA19-9 carrier specific to pancreatic cancer.

Among the four proteins selected from the MS identification, three proteins were failed to be confirmed as CA19-9 carriers in the follow-up serum profiling assay. Compared to the variation results from the instrument and the sample processing, the failure is more likely due to the selection of candidate sample pool. CA19-9 presents on a variety of proteins in different subpopulations of pancreatic cancer patients. The carriers identified from a pool of 5-10 patients do not necessarily represent a bigger population. In the case of ApoB, multiple peptides were identified in all three experiments, suggesting a high prevalence of ApoB as carrier of CA19-9, which was confirmed by the follow-up microarray. On the contrary, ARVCF was only detected by MS for several peptides in one cancer pool. In our profiling of the new sample set with 124 cancer patients, slight signal of CA19-9 on ARVCF was only found in 3 of 124 patients(Figure 18), indicating a much less frequency of ARVCF as a CA19-9 carrier. Suggested by these findings, in our future study to identify more CA19-9 carriers, we will 1) raise the selection threshold of candidates from MS identification for the follow-up microarray assay; and 2) further stratify target patients to identify the niche for applying a particular CA19-9 carrier as biomarker.

Another level of stratification is also needed for pooling the selected false-positives for MS identification. The current design pooled all false positives defined by total CA19-9 in a sample set regardless of their actual CA19-9 levels. However, the false-positives with extremely high level of CA19-9 may indicate a hidden association of pancreatic cancer that is not yet identified. Thus, comparing the carriers from these samples with true cancer samples by searching for the difference may fail to identify any hits specific to cancer. A better strategy is to divide the false positives to two cohorts according to their level of total CA19-9 and compare each cohort with true cancer.

The entire family of mucins was not targeted in the MS identification due to the lack of trypsin digestion site in the characteristic tandem repeat region of this protein family. Given to the heavy glycosylation and their tight association to cancer, a general profiling of CA19-9 on all other mucins for which antibodies are available may be worth to implement to find the CA19-9 elevation related to pancreatic cancer. Meanwhile, in order to fully profile the CA19-9 level on the protein, it is better to use more than one antibody (if available) in the antibody sandwich assays, especially for the protein targets which are highly glycosylated. Heavy-glycosylation of a protein likely suggests a higher susceptibility of its performance affected by the selection of antibodies, which may prefer different glycoforms of the protein [119].

In summary, this work was inspired by the diversity of CA19-9 distribution on different disease states, and presented a pilot exploration in identifying novel CA19-9 carrier proteins in pancreatic cancer patients. ApoB is the first CA19-9 carrier protein identified from pancreatic cancer patients but with CA19-9 signal not related to the disease. It represents a yet-to-be-identified group of CA19-9 carriers which do not specifically work for the development of pancreatic cancer and may provide interfering CA19-9 signals in some non-cancer patients or controls. Better strategies have been derived from this study and will be contributed to the future attempts of discovering CA19-9 carriers. The CA19-9 carriers identified using this workflow in the future study may contribute to two applications. One is to differentiate from the false positive and true positive, the other is to pick up the false negatives which have decent CA19-9 signals but not sufficient to be differentiated from non-cancer groups.

5.5 Materials and methods

Serum and plasma samples. Serum samples from pancreatic cancer, pancreatitis and healthy subjects were collected at Evanston Northwestern Healthcare (ENH), University of Pittsburgh (UP) School of Medicine. All sample collection was under protocols approved by a local Institutional Review Board. The healthy control samples from ENH were collected from high-risk individuals from pancreatic-cancer-prone families undergoing surveillance with Endoscopic Ultrasound (EUS) or Endoscopic Retrograde Cholangiopancreatography (ERCP). The control subjects had no pancreatic lesions. All samples were stored at -80°C and sent frozen on dry ice. Each aliquot had been thawed no more than three times before use.

Antibodies. The antibodies were obtained from various sources. All antibodies were screened for reactivity and integrity, purified, and prepared at 0.25 mg/ml in pH 7.2 PBS. The steps of antibody purification included ultracentrifugation at 47,000g at 4 degree for 1 hour and dialysis (Slide-A-Lyzer Mini Dialysis Units, PIERCE) against pH 7.2 PBS at 4 degree for 2 hours.

Microarray fabrication. Approximate 175 pg (350 pl at 500 µg/ml or 700 pl at 250 µg/ml) of each antibody was spotted on the surfaces of ultra-thin nitrocellulose-coated microscope slides (PATH slides; GenTel Biosciences) by a piezoelectric non-contact printer (Biochip Arrayer; PerkinElmer Life Sciences) for the slides used in ENH sample set, and by a non-contact arrayer at GenTel Biosciences for the slides used in UP set. Forty-eight identical arrays containing triplicates of all selected antibodies were printed on each slide. The three spots in a triplicate were positioned adjacent to each other in the slides used in ENH and UM sets, and the position of each spot in triplicates was randomly assigned in the slides used in UP sample set.

Hydrophobic borders were imprinted around each array using a stamping device (SlideImprinter, The Gel Company, San Francisco, CA).

Microarray assays. Microarray sandwich assays were performed to measure either the level of total CA19-9 and its level on the selected proteins captured by the immobilized antibodies (Figure 16a). The sandwich assay consisted of four 1-hour-incubations in room temperature (RT) with the following reagents: 1) blocking buffer (PBS containing 0.5% Tween-20 (PBST0.5) and 1% BSA); 2) a serum sample, diluted two-fold in 1XTBS containing 0.08% Brij, 0.08 Tween-20, 50 µg/ml protease inhibitor cocktail (Complete Protease Inhibitor Tablet, Roche Applied Science), and a cocktail of IgG from mouse, goat, and sheep each at 100 µg/ml and rabbit IgG at 200 µg/ml (Jackson ImmunoResearch Laboratories, Inc.); 3) biotinylated detection antibody or lectin (2 µg/ml), diluted in PBST0.1 containing 0.1% BSA; 4) streptavidin-phycoerythrin (10 µg/ml, Roche Applied Science), diluted in PBST0.1 containing 0.1% BSA. After each step, the slides were rinsed in three baths of PBST0.1 and dried by centrifugation (Eppendorf 5810R, rotor A-4-62, 1500 x g).

Fluorescence emission from the phycoerythrin was detected at 570 nm using a microarray scanner (LS Reloaded, Tecan). All arrays within one slide were scanned at a single laser power and detector gain setting. The images were quantified using the software program GenePix Pro 5.0 (Molecular Devices, Sunnyvale, CA). Spots were identified using automated spot-finding or manual adjustments for occasional irregularities. The local backgrounds were subtracted from the median intensity of each spot, and triplicate spots were averaged using the geometric mean. The coefficient of variation between replicate analyzed spots was 10-15%.

Immunoprecipitation. Serum samples (pools) from selected patients were incubated with biotinylated antibody against CA19-9 (1B.844, UsBiological) (1 µl serum with 2 µg

antibody) and 50 µg/ml protease inhibitor cocktail (Complete Protease Inhibitor Tablet, Roche Applied Science) at 4° C overnight with gentle rotation. Superparamagnetic beads of 2.8 µm in diameter with a streptavidin monolayer covalently coupled to the surface was obtained commercially (Dynabeads M-280 Streptavidin, Invitrogen). Dynabeads (2 µl dynabeads to 1 µg antibody) were washed with 1X PBS containing 0.1% Tween-20 and 0.1mg/ml Dexton for three times before and after the 1 hour nutation at room temperature in the same buffer. The beads were next coupled with the mix of antibody and serum for 1.5 hour at room temperature under gentle rotation. After the completion of coupling, the beads were washed with 1XPBS with 0.1% Tween-20 for three times at room temperature. For western blotting, the antigens were eluted by being boiled with 1.5X SDS gel loading buffer (75 mM Tris, pH 6.8; 3% SDS; 15% Glycerol; 3.75% b-Mercaptoethanol; 0.03% Bromophenol blue) at 100° C for 5 min. For deglycosylation, the beads remained untouched after the final washed and processed as described in the deglycosylation section.

Western blot. The samples were fractionated on Bis-Tris 4-12% XT precast gels (Bio-Rad) and transferred to nitrocellulose (0.45µm, Bio-Rad). We used 5% milk in TBST0.05 for overnight blocking. Primary detection was completed by a 60 min incubation of 3 µg/ml mouse monoclonal antibodies, and secondary detection was accomplished by incubating HRP conjugated goat anti-mouse IgG (1:200,000) (ImmunoPure, Pierce Biotechnology) for another 60 min.

Deglycosylation. Deglycosylation was performed using a commercial enzymatic kit (Enzyme CarboRelease Kit, QAbio). In brief, washed beads from immunoprecipitation was first incubated with reaction buffer and denaturant from the kit and boiled for 5 min at 100° C. After chilled on ice, triton-X and enzymes were added to the beads (1 µl of each enzyme for 20 µl

serum immunoprecipitated) followed by a 4.5 hours incubation at 37° C. At the end, 1.5X SDS gel loading buffer (75 mM Tris, pH 6.8; 3% SDS; 15% Glycerol; 3.75% b-Mercaptoethanol; 0.03% Bromophenol blue) (1 µl buffer for 2 µl initial beads slurry) were added and samples were boiled at 100° C for 5 min to ensure fully elution of proteins from the beads.

Mass spectrometry analysis. Deglycosylated samples with SDS were initially separated by migrating into Bis-Tris 4-12% XT precast gel (Bio-Rad) for approximately 2cm. The entire gel was fixed for 2 hours at room temperature in the solution containing 40% methanol, 10% acetic acid and 50% water. Mass spectrum identification was performed by Mary Hurley at Michigan Proteome Consortium (Ann Arbor, MI). For each sample, two gel slices were obtained. One for proteins above 250kDa, the other contains proteins sizing between 75kDa to 250kDa. The samples were digested in-gel trypsin and were analyzed by capillary reversed-phase HPLC coupled offline to a MALDI TOFTOF tandem mass spectrometer (ABI model 4800). The proteins were identified by searching a genome database using Protein Pilot software.

Statistical analysis. The microarray data were sorted for each capture antibodies by SAS. The area-under-the-curves (AUC) were calculated from the receiver-operator characteristic (ROC) analysis, the Mann-Whitney rank-sum tests and the box plots were performed using OriginPro 8. Clustering and visualization were performed using the programs Cluster and Treeview and MultiExperiment Viewer.

APPENDICES

TABLES

Type	Exp #	N	Unused	Total	%Cov(95)	Accession	Name	Peptides(95%)
Late cancer pool	Exp#3	5	7.8	7.8	17.67	IPI:PII00021842.1	APOE Apolipoprotein E	4
Benign pool	Exp#3	8	1.1	1.1	0.00	IPI:PII00021842.1	APOE Apolipoprotein E	0
Healthy pool	Exp#3	10	3.1	3.1	4.73	IPI:PII00021842.1	APOE Apolipoprotein E	1
Late cancer pool	Exp#2	7	3	3	1	IPI:PII00879055.1	ARVCF ARVCF protein	1
Early cancer pool	Exp#2	18	0	2	1	IPI:PII00879055.1	ARVCF ARVCF protein	1
Benign or healthy	Exp#2			Not detected			N/A	
Late cancer pool	Exp#2	13	0	2	1	IPI:PII00032328.2KNG1	Isoform HMW of Kininogen-1	1
Early cancer pool	Exp#2	8	0	7	3	IPI:PII00032328.2KNG1	Isoform HMW of Kininogen-1	3
Benign pool	Exp#2	12	0	2	0	IPI:PII00032328.2KNG1	Isoform HMW of Kininogen-1	0
Healthy pool	Exp#2	8	0	3	1	IPI:PII00032328.2KNG1	Isoform HMW of Kininogen-1	1
Late cancer pool	Exp#2	13	0	2	2	IPI:PII00215894.1KNG1	Isoform LMW of Kininogen-1	1
Early cancer pool	Exp#2	8	7	7	4	IPI:PII00215894.1KNG1	Isoform LMW of Kininogen-1	3
Benign pool	Exp#2	12	2	2	0	IPI:PII00215894.1KNG1	Isoform LMW of Kininogen-1	0
Healthy pool	Exp#2	13	0	2	2	IPI:PII00215894.1KNG1	Isoform LMW of Kininogen-1	1

Footnotes

N: The rank of the specified protein relative to all other proteins in the list of detected proteins.

Unused: (ProtScore) A measure of the protein confidence for a detected protein, calculated from the peptide confidence for peptides from spectra that have not already been completely “used” by higher scoring winning proteins.

Total: (ProtScore): A measure of the total amount of evidence for a detected protein. The total ProtScore is calculated using all of the peptides detected for the protein.

%Cov (95): The number of matching amino acids (from peptides) divided by the total number of amino acids in the sequence, expressed as a percentage.

Accession: The accession number for the detected protein.

Name: The name of the detected protein.

Table 4 Mass spectrometry identification of CA19-9 carriers.

Sample	Exp #	N	Unused	Total	%Cov (95)	Accession	Name	Peptides (95%)
Late Cancer	Exp#1	1	142	142	20	IPI:IP100022229.1	APOB Apolipoprotein B-100	67
	Exp#2	3	21	21	2	IPI:IP100022229.1	APOB Apolipoprotein B-100	10
	Exp#3	1	91.4	91.4	12.01	IPI:IP100022229.1	APOB Apolipoprotein B-100	45
Early Cancer	Exp#1	1	126	126	18	IPI:IP100022229.1	APOB Apolipoprotein B-100	62
	Exp#2	2	31	31	3	IPI:IP100022229.1	APOB Apolipoprotein B-100	13
	Exp#3	4	4.4	4.4	0.50	IPI:IP100022229.1	APOB Apolipoprotein B-100	2
Pancreatitis	Exp#1	3	3	3	1	IPI:IP100022229.1	APOB Apolipoprotein B-100	2
	Exp#2	1	43	43	5	IPI:IP100022229.1	APOB Apolipoprotein B-100	19
	Exp#3	1	78.5	78.5	9.34	IPI:IP100022229.1	APOB Apolipoprotein B-100	36
Healthy	Exp#2	5	13	13	1	IPI:IP100022229.1	APOB Apolipoprotein B-100	6
	Exp#3	1	83.3	83.3	10.56	IPI:IP100022229.1	APOB Apolipoprotein B-100	38

Footnotes

N: The rank of the specified protein relative to all other proteins in the list of detected proteins.

Unused: (ProtScore) A measure of the protein confidence for a detected protein, calculated from the peptide confidence for peptides from spectra that have not already been completely “used” by higher scoring winning proteins.

Total: (ProtScore): A measure of the total amount of evidence for a detected protein. The total ProtScore is calculated using all of the peptides detected for the protein.

%Cov (95): The number of matching amino acids (from peptides) divided by the total number of amino acids in the sequence, expressed as a percentage.

Accession: The accession number for the detected protein.

Name: The name of the detected protein.

Peptides (95%): The number of peptide hits at 95% confidence.

Table 5 Mass spectrometry identification of apolipoprotein B-100.

FIGURES

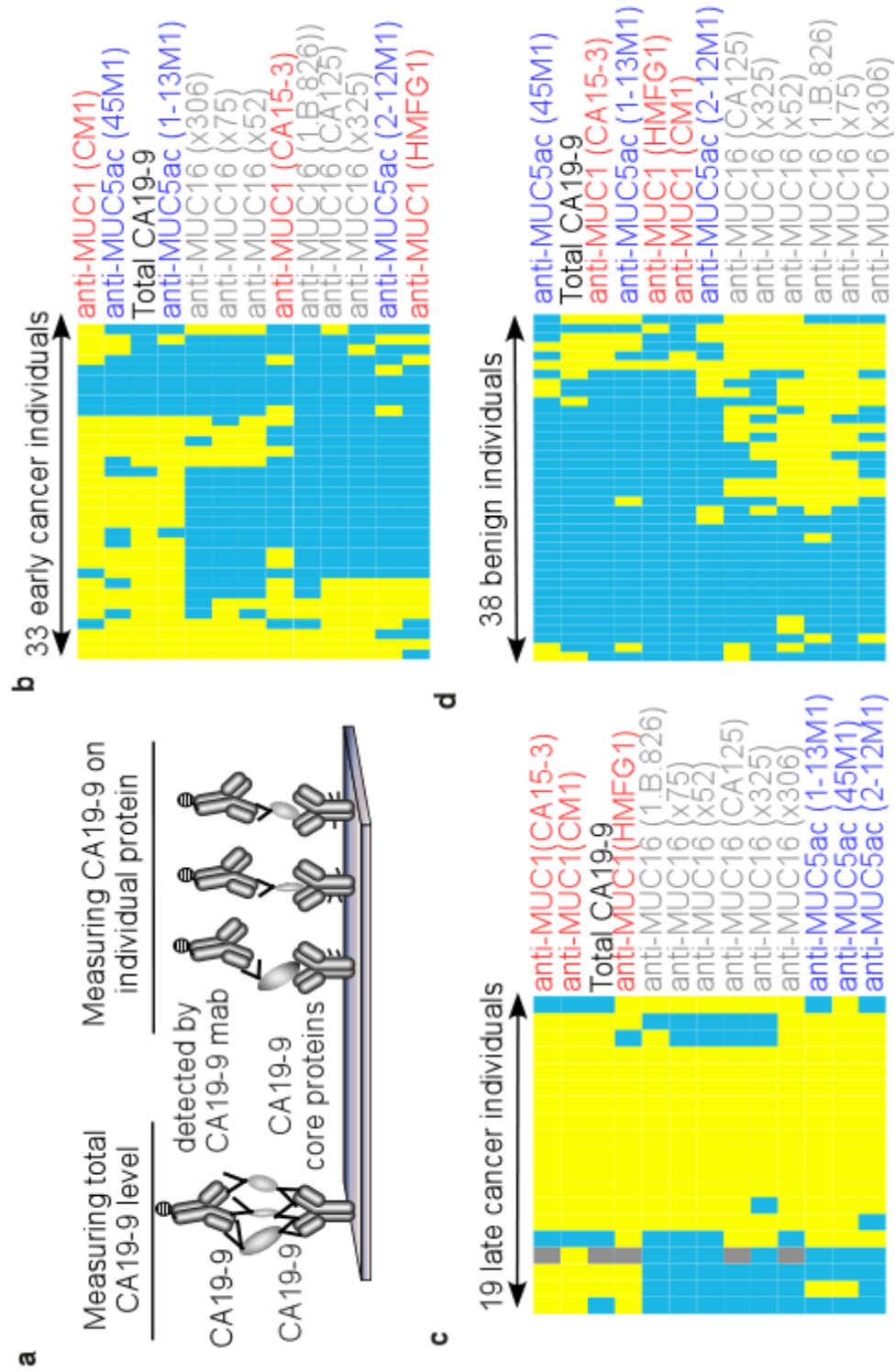


Figure 16 Detection of CA19-9 on pre-identified protein MUC1, MUC5ac, MUC16.

a, Detection of CA19-9 by microarray for its total level (left) and its level on individual proteins (right). Monoclonal antibodies targeting CA19-9 and individual proteins were immobilized on microarray slide and detected by CA19-9 antibody. b-d, Level of CA19-9 measured on MUC1, MUC5ac and MUC16 from early stage-(b), late stage-(c), and benign (d) patients. Captured antibodies are indicated by rows with clone number in parenthesis. Yellow, high level of CA19-9; Blue, low CA19-9 level.

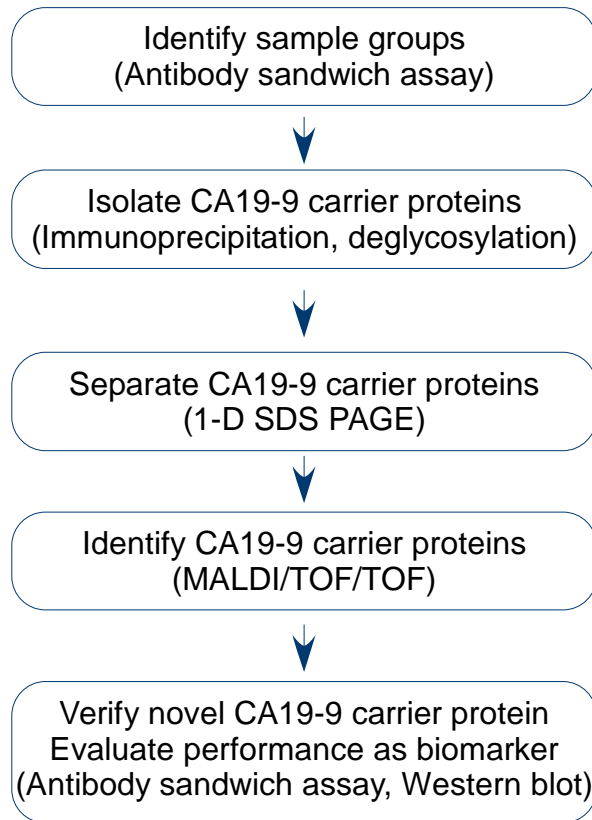


Figure 17 Schematic of experiment design.

Parenthesis indicates methods used for the purpose stated previously.

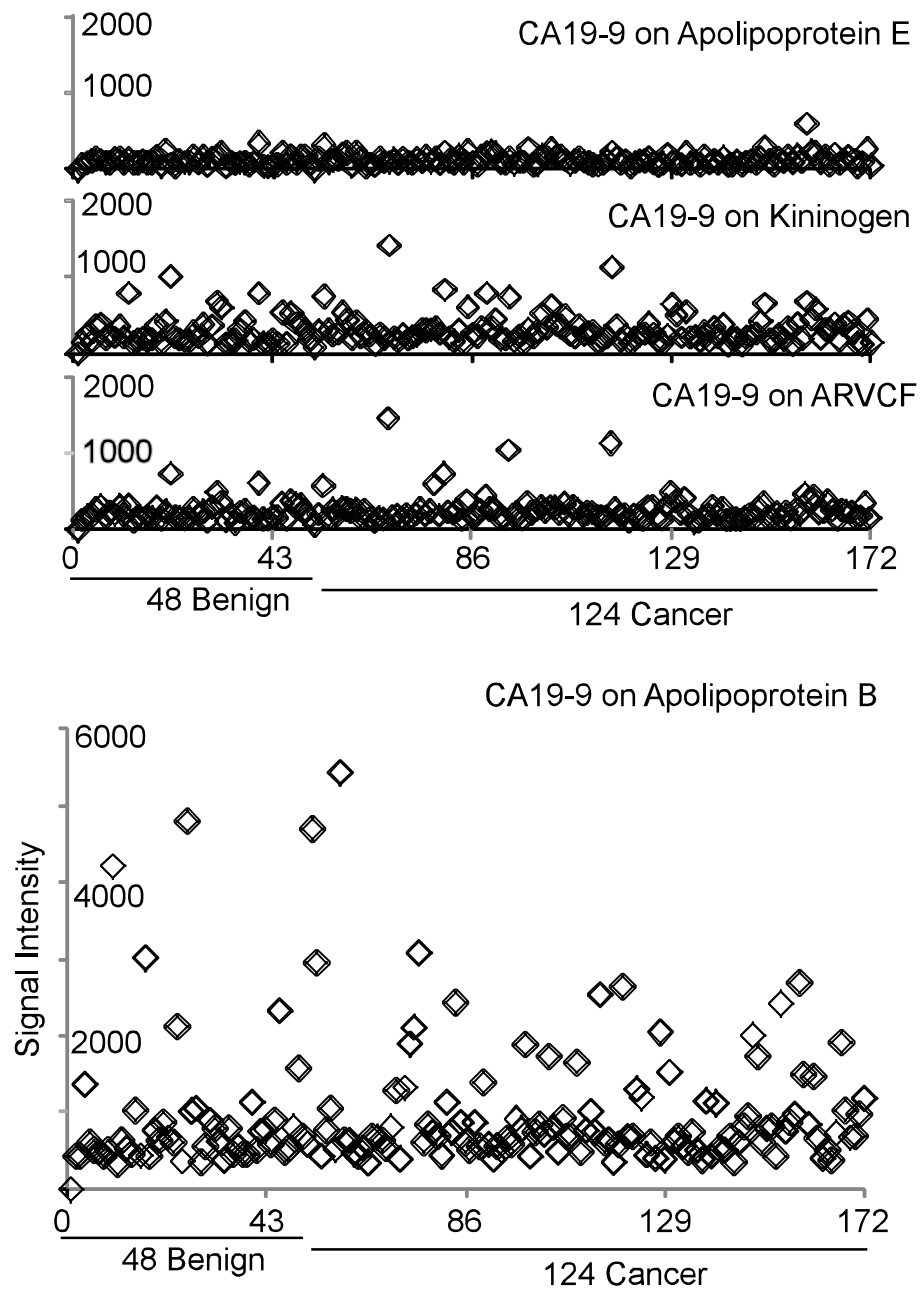


Figure 18 Detection of CA19-9 on selected proteins.

CA19-9 on Apo B (left) and three other candidate proteins Aplipoprotein E, Kininogen, and ARVCF (right) from 124 cancer patients and 48 pancreatitis (benign) patients were shown.

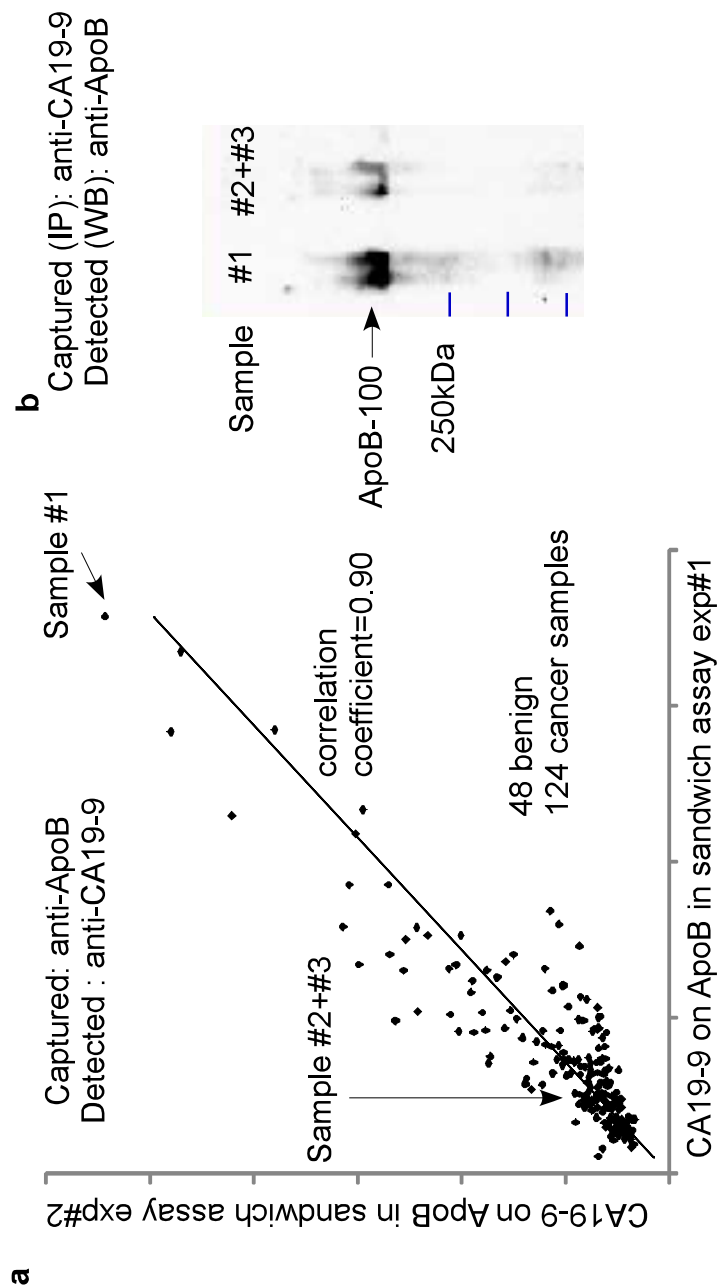


Figure 19 Detection of Apo B as CA19-9 carrier.

a, correlation of CA19-9 signal on ApoB between duplicated antibody sandwich assays. ApoB was captured by antibody immobilized on the microscopic slide and detected by CA19-9 antibody. b, validation of CA19-9 signal on ApoB by immunoprecipitation using CA19-9 antibody and blotting with antibody targeting ApoB using three samples identified from a.

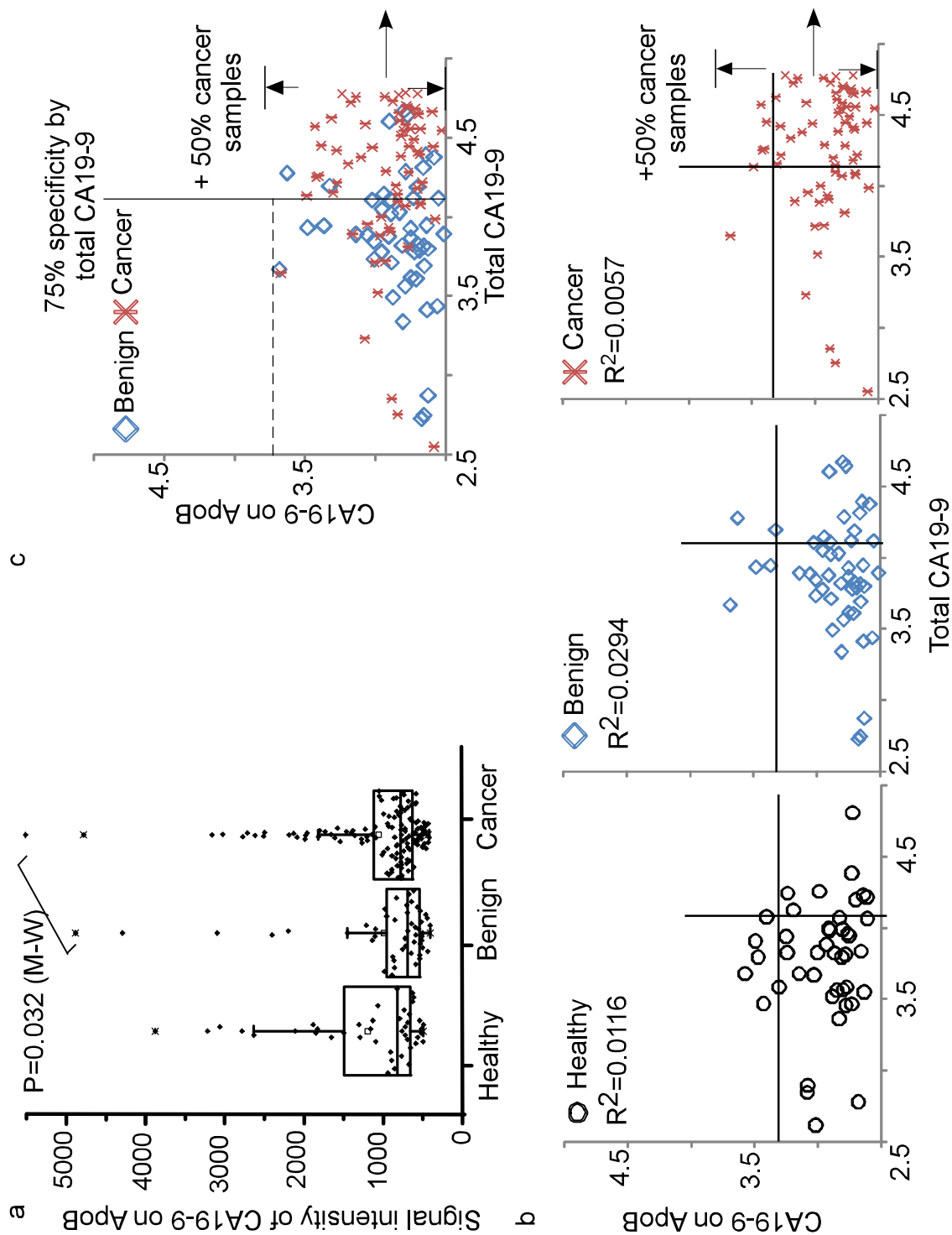


Figure 20 Apo B as CA19-9 carrier in healthy, pancreatitis and pancreatic cancer groups.

Apo B as CA19-9 carrier in healthy, pancreatitis and pancreatic cancer groups. a, Box-plot of CA19-9 level on ApoB from 43 healthy, 48 pancreatitis (pancreatitis) and 124 cancer

individuals. Each point represent an individual sample. The boxes indicate the quartiles, with the median indicated by the solid horizontal lines and the ranges marked by the vertical lines. P-value, Mann-Whitney test. b. Comparison of CA19-9 on ApoB and total CA19-9 in healthy subjects (left), pancreatitis group (middle) and cancer group (right). Solid lines indicate threshold selected for the respective dimension. For CA19-9, the threshold determines the 75% specificity, for CA19-9 on ApoB, the threshold was selected from Figure 18.. c, Scatter plot of CA19-9 on ApoB (y-axis) and total CA19-9 (x-axis). Each point represents a sample from 124 pancreatic cancer or 48 pancreatitis patients. The cancer patients with total CA19-9 signal in the higher half were omitted in the figure for concision purpose, but the range of CA19-9 signal on ApoB from the 50% patients were marked by the arrows. R square, coefficient of determination.

CHAPTER 6 SYNTHESIS AND FUTURE WORK

This dissertation work was aiming at developing biomarker for pancreatic cancer detection using the alteration of glycosylation on serum mucins. With the aid of glycoproteomic tools such as antibody microarray and sandwich assays using lectin detection, we were able to profile the glycosylation changes on three mucins, MUC1, MUC5ac, and MUC16 (Chapter 3). Suggested by the glycan profiling on the three mucins, we focused on the established pancreatic cancer glycan marker CA19-9, and profiled its changes on multiple proteins. This led us to an enhanced detection of pancreatic cancer by combining different derivatives of the standard marker CA19-9 (Chapter 4). To further understand the CA19-9 marker, we searched for the potential new protein carriers of this glycan described in Chapter 5. The novel carrier identified in this chapter did not contribute to the CA19-9 elevation in pancreatic cancer, providing us an example explaining the limitation of using standard marker CA19-9 (Chapter 5).

6.1 Monitoring cancer-associated glycan alteration on selected proteins

Alterations of glycans or glycoproteins were monitored clinically to control cancer occurring in pancreas, colon, breast, prostate, and ovarian [14]. There are two major ways that the glycosylation may be altered in cancer. One is through the production of tumor-specific glycans, regardless of what proteins they rely on. Thus, capturing the change of these glycans enables the detection of the cancer. Alternatively, tumor may express abnormal amount of glycoproteins to result in a change of glycosylation. Researchers in the past decades have been intensively focused on either of the two sides and have developed cancer markers such as glycan CA19-9 for pancreatic cancer and glycoprotein CA125 for ovarian cancer. With the identification of the diversity of targeted patients, the clinical usefulness of these markers is

limited due to the existence of false-negatives and false-positives. Therefore, reducing the false rate of current glycosylation markers is essential in enhancing their clinical usefulness.

In this dissertation work, we have demonstrated that, rather than exploring glycan or glycoprotein separately, investigating the relationship between them improved the detection. By profiling various glycosylation on a particular glycoprotein (Chapter 3) or examining the prevalence of a specific glycan on multiple glycoproteins (Chapter 4), we have achieved a more accurate discrimination over using either the glycoprotein or the glycan alone.

We profiled the alteration at both protein level and total glycan level for three mucins in the serum from pancreatic cancer patients and healthy controls (Chapter 3). As a result, we found only mild elevation at glycoprotein level for all three mucins, but more distinct elevation of their glycosylations. The increased glycosylation fell into 6 major groups including both N- and O-link glycans, both elongated and cancer-specific truncated forms (TF glycan and Tn antigen), as well as the Lewis antigens. However, all of these were also well observed from other types of cancer [17] or non-cancerous pancreatic lesions [144]. From a biomarker perspective, the changes on protein level would have insufficient sensitivity, and the change on glycan level may raise concern for specificity. Therefore, we monitored the relationship of glycan changes on individual protein. The overall-elevated TF glycan was found highly increased on MUC5ac, slightly increased on MUC1, and decreased on MUC16. So is the Tn antigen, which increased on MUC1 and MUC5ac, but decreased on MUC16. Compared to the change at protein level for MUC5ac, the additional elevation of the TF or Tn glycan enhanced the detection of cancer patients. On the other hand, compared to the change of a glycan which may be common to multiple conditions, the various patterns of its alteration on selected proteins may provide a unique indication to a particular disease and can be potentially used to assist the specificity.

Inspired by the above findings, we developed biomarker panel using derivatives of CA19-9 marker to accomplish enhanced detection of pancreatic cancer (Chapter 4). Profiling CA19-9 changes on MUC1, MUC5ac, and MUC16 detected a total of 25% to 100% of false-negative samples labeled by total CA19-9 without reduction of specificity. Indeed, we found that the newly identified CA19-9 carrier ApoB did not contribute to the CA19-9 elevation on pancreatic cancer, explaining why total CA19-9 detection of pancreatic cancer can be improved by the measurement of CA19-9 on cancer-specific carriers. Altogether, these findings highlight the value of improving the molecular resolution of a glycan marker: in addition to its total level, its prevalence on specific proteins should also be examined.

6.2 Biological and/or statistical selection of biomarker panel

The standard way of selecting biomarkers into a panel is by investigating the overall performance of each combination of biomarkers to return with a panel performing the best. This results in hidden issues in practice, especially when statistics is not speaking for biology. An example arose from the analysis of datasets generated from a double-blinded biomarker study described in Chapter 4, We screened a total of 70 candidate markers generated by applying 5 glycan detections on proteins captured by 14 different antibodies. The best 5-marker panel selected by forward logistic regression excluded the standard marker CA19-9 and included a marker with only poor elevation on a few samples. However, when each marker was examined individually, elevation of total CA19-9 was one of the dominant changes suggested by the data. Furthermore, there was no sufficient proof showing that this marker can be replaced, though it was suggested by the statistically selected panel.

Instead of being led by statistics, we selected the panel by exploring the biological complementarity between markers. Subgroups of patients were first identified by the dominant marker CA19-9. Next, additional markers were specifically selected to target the false-negatives without affecting the true-negatives. The major advantage of this selection strategy is that it takes into consideration the natural existence of subgroups and stratifies patients based on that. Compared to a panel statistically selected, in which case each marker is evaluated for their performance in the whole sample set, the panel selected by complementarity is more efficient in the sense that each additional marker is only evaluated in a pre-stratified patient group determined by the initial marker.

Both strategies have their own pros and cons. Statistical strategy features for its powerfulness but limited in the ignorance of biological details. Statistical selection is particularly useful in providing a general evaluation of multiple markers when none of them is dominant. The complementarity strategy is the reverse. This strategy relies on the stratification of patients and therefore works the best for a disease with the dominant marker(s) available. An effective working strategy for future marker selection should incorporate both, using the statistical strategy to provide guidelines and identify candidate dominant markers, and then applying the complementarity strategy based on the dominant markers to customize the further selection panel to reveal particular biological details.

6.3 Future work and conclusion remark

This dissertation demonstrates the effectiveness of a new platform for development of glycan biomarkers in pancreatic cancer. The platform includes 1) a technological solution using glycoproteomics methods based on antibody microarray, and 2) an analytical solution to identify,

evaluate, select and apply these biomarkers. The ultimate goal of the study is to improve the clinical control of pancreatic cancer. One of the immediate needs, as discussed in Chapter 4, is to assist physicians with simple serum assay to exclude pancreatitis patients from the decision of invasive diagnosis for cancer. To accomplish this goal, based on the platform and progress already established in the dissertation project, future work should work with larger sample sets and focus on identifying biomarkers to target the false negative cancer patients that can't be detected by the panel in Chapter 4. Two major subgroups account for the remaining false negatives and they require distinct solutions. For the subgroup of false negatives with CA19-9 presented, identifying the CA19-9 carriers in the group, and using the measurement of CA19-9 on these carrier proteins as biomarkers is the preferred solution. On the other hand, to target the group without total CA19-9 signals, identifying new glycan markers should be considered. The altered glycosylations on the three mucins described in Chapter 3 is a promising pool of candidate markers for this purpose. Finally, careful evaluation of these markers using both statistical and complementarity strategy will advance our understanding of the biomarker performance, and ultimately enhance our control of this disastrous disease. The technological platform and analytical strategy demonstrated in this dissertation are also useful for the biomarker evaluation in other clinical settings (e.g. monitoring therapy response) as well as for the glycan biomarker development in a different disease.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Jemal, A., et al., *Cancer Statistics, 2010*. CA Cancer J Clin, 2010.
2. *Guidelines for the management of patients with pancreatic cancer perampullary and ampullary carcinomas*. Gut, 2005. **54 Suppl 5**: p. v1-16.
3. Wong, H.H. and N.R. Lemoine, *Pancreatic cancer: molecular pathogenesis and new therapeutic targets*. Nat Rev Gastroenterol Hepatol, 2009. **6**(7): p. 412-22.
4. Hruban, R.H., A. Maitra, and M. Goggins, *Update on pancreatic intraepithelial neoplasia*. Int J Clin Exp Pathol, 2008. **1**(4): p. 306-16.
5. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**(1): p. 57-70.
6. Miura, F., et al., *Diagnosis of pancreatic cancer*. HPB (Oxford), 2006. **8**(5): p. 337-42.
7. Karmazanovsky, G., et al., *Pancreatic head cancer: accuracy of CT in determination of resectability*. Abdom Imaging, 2005. **30**(4): p. 488-500.
8. Pannu, H.K. and E.K. Fishman, *Complications of endoscopic retrograde cholangiopancreatography: spectrum of abnormalities demonstrated with CT*. Radiographics, 2001. **21**(6): p. 1441-53.
9. Herrero-Zabaleta, M.E., et al., *Monoclonal antibody against sialylated Lewis(a) antigen*. Bull Cancer, 1987. **74**(4): p. 387-96.
10. Berger, A.C., et al., *Postresection CA 19-9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: a prospective validation by RTOG 9704*. J Clin Oncol, 2008. **26**(36): p. 5918-22.
11. Hess, V., et al., *CA 19-9 tumour-marker response to chemotherapy in patients with advanced pancreatic cancer enrolled in a randomised controlled trial*. Lancet Oncol, 2008. **9**(2): p. 132-8.
12. Hidalgo, M., *Pancreatic cancer*. N Engl J Med, 2010. **362**(17): p. 1605-17.

13. Harsha, H.C., et al., *A compendium of potential biomarkers of pancreatic cancer*. PLoS Med, 2009. **6**(4): p. e1000046.
14. Dube, D.H. and C.R. Bertozzi, *Glycans in cancer and inflammation--potential for therapeutics and diagnostics*. Nat Rev Drug Discov, 2005. **4**(6): p. 477-88.
15. Varki, A., et al., *Essentials of glycobiology*. 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. xvii, 653 p.
16. Dennis, J.W., M. Granovsky, and C.E. Warren, *Glycoprotein glycosylation and cancer progression*. Biochim Biophys Acta, 1999. **1473**(1): p. 21-34.
17. Fuster, M.M. and J.D. Esko, *The sweet and sour of cancer: glycans as novel therapeutic targets*. Nat Rev Cancer, 2005. **5**(7): p. 526-42.
18. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
19. Hirabayashi, J., et al., *Oligosaccharide specificity of galectins: a search by frontal affinity chromatography*. Biochim Biophys Acta, 2002. **1572**(2-3): p. 232-54.
20. Culf, A.S., M. Cuperlovic-Culf, and R.J. Ouellette, *Carbohydrate microarrays: survey of fabrication techniques*. Omics, 2006. **10**(3): p. 289-310.
21. Wang, D., *Carbohydrate microarrays*. Proteomics, 2003. **3**(11): p. 2167-75.
22. Houseman, B.T. and M. Mrksich, *Carbohydrate arrays for the evaluation of protein binding and enzymatic modification*. Chem Biol, 2002. **9**(4): p. 443-54.
23. Park, S. and I. Shin, *Fabrication of carbohydrate chips for studying protein-carbohydrate interactions*. Angew Chem Int Ed Engl, 2002. **41**(17): p. 3180-2.
24. Fazio, F., et al., *Synthesis of sugar arrays in microtiter plate*. J Am Chem Soc, 2002. **124**(48): p. 14397-402.
25. Wang, D., et al., *Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells*. Nat Biotechnol, 2002. **20**(3): p. 275-81.

26. Fukui, S., et al., *Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions*. Nat Biotechnol, 2002. **20**(10): p. 1011-7.
27. Willats, W.G., et al., *Sugar-coated microarrays: a novel slide surface for the high-throughput analysis of glycans*. Proteomics, 2002. **2**(12): p. 1666-71.
28. Xia, B., et al., *Versatile fluorescent derivatization of glycans for glycomic analysis*. Nat Methods, 2005. **2**(11): p. 845-50.
29. Blixt, O., et al., *Printed covalent glycan array for ligand profiling of diverse glycan binding proteins*. Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17033-8.
30. Rudiger, H. and H.J. Gabius, *Plant lectins: occurrence, biochemistry, functions and applications*. Glycoconj J, 2001. **18**(8): p. 589-613.
31. Manimala, J.C., et al., *High-throughput carbohydrate microarray analysis of 24 lectins*. Angew Chem Int Ed Engl, 2006. **45**(22): p. 3607-10.
32. Manimala, J.C., et al., *High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems*. Glycobiology, 2007. **17**(8): p. 17C-23C.
33. Manimala, J.C., et al., *Carbohydrate array analysis of anti-Tn antibodies and lectins reveals unexpected specificities: implications for diagnostic and vaccine development*. Chembiochem, 2005. **6**(12): p. 2229-41.
34. Moller, I., et al., *High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles*. Glycoconj J, 2008. **25**(1): p. 37-48.
35. Huang, C.Y., et al., *Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen*. Proc Natl Acad Sci U S A, 2006. **103**(1): p. 15-20.
36. Lawrie, C.H., et al., *Cancer-associated carbohydrate identification in Hodgkin's lymphoma by carbohydrate array profiling*. Int J Cancer, 2006. **118**(12): p. 3161-6.

37. Wang, C.C., et al., *Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer*. Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11661-6.
38. Wang, D., et al., *Photogenerated glycan arrays identify immunogenic sugar moieties of Bacillus anthracis exosporium*. Proteomics, 2007. **7**(2): p. 180-4.
39. Blixt, O., et al., *Pathogen specific carbohydrate antigen microarrays: a chip for detection of Salmonella O-antigen specific antibodies*. Glycoconj J, 2008. **25**(1): p. 27-36.
40. Stevens, J., et al., *Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities*. J Mol Biol, 2006. **355**(5): p. 1143-55.
41. Disney, M.D. and P.H. Seeberger, *The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens*. Chem Biol, 2004. **11**(12): p. 1701-7.
42. Park, S., et al., *Carbohydrate chips for studying high-throughput carbohydrate-protein interactions*. J Am Chem Soc, 2004. **126**(15): p. 4812-9.
43. Park, S. and I. Shin, *Carbohydrate microarrays for assaying galactosyltransferase activity*. Org Lett, 2007. **9**(9): p. 1675-8.
44. Blixt, O., et al., *Glycan microarrays for screening sialyltransferase specificities*. Glycoconj J, 2008. **25**(1): p. 59-68.
45. Sharon, N. and H. Lis, *History of lectins: from hemagglutinins to biological recognition molecules*. Glycobiology, 2004. **14**(11): p. 53R-62R.
46. Sharon, N. and H. Lis, *Lectins as cell recognition molecules*. Science, 1989. **246**(4927): p. 227-34.
47. Sharon, N. and H. Lis, *The structural basis for carbohydrate recognition by lectins*. Adv Exp Med Biol, 2001. **491**: p. 1-16.
48. Varki, A., et al., *Essentials of Glycobiology*. 1999: Cold Spring Harbor Laboratory Press.

49. Sharon, N., *Lectins: carbohydrate-specific reagents and biological recognition molecules*. J Biol Chem, 2007. **282**(5): p. 2753-64.
50. Osako, M., et al., *Immunohistochemical study of mucin carbohydrates and core proteins in human pancreatic tumors*. Cancer, 1993. **71**(7): p. 2191-9.
51. Satomura, Y., et al., *Expression of various sialylated carbohydrate antigens in malignant and nonmalignant pancreatic tissues*. Pancreas, 1991. **6**(4): p. 448-58.
52. Shimizu, K., et al., *Comparison of carbohydrate structures of serum alpha-fetoprotein by sequential glycosidase digestion and lectin affinity electrophoresis*. Clin Chim Acta, 1996. **254**(1): p. 23-40.
53. Thompson, S., et al., *Abnormally-fucosylated haptoglobin: a cancer marker for tumour burden but not gross liver metastasis*. Br J Cancer, 1991. **64**(2): p. 386-90.
54. Okuyama, N., et al., *Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation*. Int J Cancer, 2005.
55. van Dijk, W., E.C. Havenaar, and E.C. Brinkman-van der Linden, *Alpha 1-acid glycoprotein (orosomucoid): pathophysiological changes in glycosylation in relation to its function*. Glycoconj J, 1995. **12**(3): p. 227-33.
56. Thompson, S., D. Guthrie, and G.A. Turner, *Fucosylated forms of alpha-1-antitrypsin that predict unresponsiveness to chemotherapy in ovarian cancer*. Br J Cancer, 1988. **58**(5): p. 589-93.
57. Hirabayashi, J., *Lectin-based structural glycomics: glycoproteomics and glycan profiling*. Glycoconj J, 2004. **21**(1-2): p. 35-40.
58. Kuno, A., et al., *Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling*. Nat Methods, 2005. **2**(11): p. 851-6.
59. Uchiyama, N., et al., *Optimization of evanescent-field fluorescence-assisted lectin microarray for high-sensitivity detection of monovalent oligosaccharides and glycoproteins*. Proteomics, 2008. **8**(15): p. 3042-50.

60. Pilobello, K.T., et al., *Development of a lectin microarray for the rapid analysis of protein glycopatterns*. *Chembiochem*, 2005. **6**(6): p. 985-9.
61. Angeloni, S., et al., *Glycoprofiling with micro-arrays of glycoconjugates and lectins*. *Glycobiology*, 2005. **15**(1): p. 31-41.
62. Koshi, Y., et al., *A fluorescent lectin array using supramolecular hydrogel for simple detection and pattern profiling for various glycoconjugates*. *J Am Chem Soc*, 2006. **128**(32): p. 10413-22.
63. Pilobello, K.T., D.E. Slawek, and L.K. Mahal, *A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome*. *Proc Natl Acad Sci U S A*, 2007. **104**(28): p. 11534-9.
64. Ebe, Y., et al., *Application of lectin microarray to crude samples: differential glycan profiling of lec mutants*. *J Biochem*, 2006. **139**(3): p. 323-7.
65. Zheng, T., D. Peelen, and L.M. Smith, *Lectin arrays for profiling cell surface carbohydrate expression*. *J Am Chem Soc*, 2005. **127**(28): p. 9982-3.
66. Chen, S., et al., *Analysis of cell surface carbohydrate expression patterns in normal and tumorigenic human breast cell lines using lectin arrays*. *Anal Chem*, 2007. **79**(15): p. 5698-702.
67. Hsu, K.L., K.T. Pilobello, and L.K. Mahal, *Analyzing the dynamic bacterial glycome with a lectin microarray approach*. *Nat Chem Biol*, 2006. **2**(3): p. 153-7.
68. Tateno, H., et al., *A novel strategy for mammalian cell surface glycome profiling using lectin microarray*. *Glycobiology*, 2007. **17**(10): p. 1138-46.
69. Tao, S.C., et al., *Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers*. *Glycobiology*, 2008. **18**(10): p. 761-9.
70. Haab, B.B., M.J. Dunham, and P.O. Brown, *Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions*. *Genome Biol*, 2001. **2**(2): p. RESEARCH0004.

71. Kjeldsen, T., et al., *Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2----6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope*. Cancer Res, 1988. **48**(8): p. 2214-20.
72. Hanisch, F.G., C. Hanski, and A. Hasegawa, *Sialyl Lewis(x) antigen as defined by monoclonal antibody AM-3 is a marker of dysplasia in the colonic adenoma-carcinoma sequence*. Cancer Res, 1992. **52**(11): p. 3138-44.
73. Thompson, S., R. Stappenbeck, and G.A. Turner, *A multiwell lectin-binding assay using lotus tetragonolobus for measuring different glycosylated forms of haptoglobin*. Clin Chim Acta, 1989. **180**(3): p. 277-84.
74. Chen, S., et al., *Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays*. Nat Methods, 2007. **4**(5): p. 437-44.
75. Forrester, S., et al., *Low-volume, high-throughput sandwich immunoassays for profiling plasma proteins in mice: identification of early-stage systemic inflammation in a mouse model of intestinal cancer*. Molecular Oncology, 2007: p. in press.
76. Orzechowski, R., et al., *Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer*. Cancer Res, 2005. **65**(23): p. 11193-202.
77. Gao, W.M., et al., *Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis*. BMC Cancer, 2005. **5**: p. 110.
78. Patwa, T.H., et al., *Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection*. Anal Chem, 2006. **78**(18): p. 6411-21.
79. Zhao, J., et al., *Glycoprotein microarrays with multi-lectin detection: unique lectin binding patterns as a tool for classifying normal, chronic pancreatitis and pancreatic cancer sera*. J Proteome Res, 2007. **6**(5): p. 1864-74.
80. Okuyama, N., et al., *Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation*. Int J Cancer, 2006. **118**(11): p. 2803-2808.

81. Burchell, J., et al., *Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin*. Cancer Res, 1987. **47**(20): p. 5476-82.
82. Forrester, S., et al., *Low-volume, high-throughput sandwich immunoassays for profiling plasma proteins in mice: identification of early-stage systemic inflammation in a mouse model of intestinal cancer*. Molecular Oncology, 2007. **1**: p. 216-225.
83. Blixt, O., et al., *Printed covalent glycan array for ligand profiling of diverse glycan binding proteins*. Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17033-17038.
84. Liang, P.H., et al., *Glycan arrays: biological and medical applications*. Curr Opin Chem Biol, 2008. **12**(1): p. 86-92.
85. Hollingsworth, M.A. and B.J. Swanson, *Mucins in cancer: protection and control of the cell surface*. Nat Rev Cancer, 2004. **4**(1): p. 45-60.
86. Brockhausen, I., *Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions*. EMBO Rep, 2006. **7**(6): p. 599-604.
87. Sanders, D.S. and M.A. Kerr, *Lewis blood group and CEA related antigens; coexpressed cell-cell adhesion molecules with roles in the biological progression and dissemination of tumours*. Mol Pathol, 1999. **52**(4): p. 174-8.
88. Stocks, S.C., M. Albrechtsen, and M.A. Kerr, *Expression of the CD15 differentiation antigen (3-fucosyl-N-acetyl-lactosamine, LeX) on putative neutrophil adhesion molecules CR3 and NCA-160*. Biochem J, 1990. **268**(2): p. 275-80.
89. Aoyagi, Y., et al., *The fucosylation index of serum alpha-fetoprotein as useful prognostic factor in patients with hepatocellular carcinoma in special reference to chronological changes*. Hepatol Res, 2002. **23**(4): p. 287.
90. Nagata, K., et al., *Mucin expression profile in pancreatic cancer and the precursor lesions*. J Hepatobiliary Pancreat Surg, 2007. **14**: p. 243-254.
91. Kim, G.E., et al., *Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas*. Gastroenterology, 2002. **123**(4): p. 1052-60.

92. Brockhausen, I., J. Schutzbach, and W. Kuhns, *Glycoproteins and their relationship to human disease*. Acta Anatomica, 1998. **161**: p. 36-78.
93. Moniaux, N., et al., *Multiple roles of mucins in pancreatic cancer, a lethal and challenging malignancy*. Br J Cancer, 2004. **91**(9): p. 1633-8.
94. Burdick, M.D., et al., *Oligosaccharides expressed on MUC1 produced by pancreatic and colon tumor cell lines*. J Biol Chem, 1997. **272**(39): p. 24198-202.
95. Lloyd, K.O., et al., *Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells*. J Biol Chem, 1996. **271**(52): p. 33325-34.
96. Storr, S.J., et al., *The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum*. Glycobiology, 2008. **18**(6): p. 456-62.
97. Springer, G.F., *T and Tn, general carcinoma autoantigens*. Science, 1984. **224**(4654): p. 1198-206.
98. Singh, R., et al., *Peanut lectin stimulates proliferation of colon cancer cells by interaction with glycosylated CD44v6 isoforms and consequential activation of c-Met and MAPK: functional implications for disease-associated glycosylation changes*. Glycobiology, 2006. **16**(7): p. 594-601.
99. Campbell, B.J., et al., *Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin*. J Clin Invest, 1995. **95**(2): p. 571-6.
100. Campbell, B.J., L.G. Yu, and J.M. Rhodes, *Altered glycosylation in inflammatory bowel disease: a possible role in cancer development*. Glycoconj J, 2001. **18**(11-12): p. 851-8.
101. Yu, L.G., et al., *Galectin-3 interaction with Thomsen-Friedenreich disaccharide on cancer-associated MUC1 causes increased cancer cell endothelial adhesion*. J Biol Chem, 2007. **282**(1): p. 773-81.
102. Magnani, J.L., et al., *Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin*. Cancer Res, 1983. **43**(11): p. 5489-92.

103. McEver, R.P., *Selectin-carbohydrate interactions during inflammation and metastasis*. Glycoconj J, 1997. **14**(5): p. 585-91.
104. Parry, S., et al., *N-Glycosylation of the MUC1 mucin in epithelial cells and secretions*. Glycobiology, 2006. **16**(7): p. 623-34.
105. Newsom-Davis, T.E., et al., *Enhanced Immune Recognition of Cryptic Glycan Markers in Human Tumors*. Cancer Res, 2009.
106. Muller, S., et al., *High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells*. J Biol Chem, 1999. **274**(26): p. 18165-72.
107. Tempero, M.A., et al., *Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer*. Cancer Res, 1987. **47**(20): p. 5501-3.
108. Kawa, S., et al., *Elevated serum levels of Dupan-2 in pancreatic cancer patients negative for Lewis blood group phenotype*. Br J Cancer, 1991. **64**(5): p. 899-902.
109. Price, M.R., et al., *Summary report on the ISOBM TD-4 Workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin*. San Diego, Calif., November 17-23, 1996. Tumour Biol, 1998. **19 Suppl 1**: p. 1-20.
110. Karsten, U., et al., *Enhanced binding of antibodies to the DTR motif of MUC1 tandem repeat peptide is mediated by site-specific glycosylation*. Cancer Res, 1998. **58**(12): p. 2541-9.
111. Burchell, J., et al., *An alpha2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas*. Glycobiology, 1999. **9**(12): p. 1307-11.
112. Lidell, M.E., J. Bara, and G.C. Hansson, *Mapping of the 45M1 epitope to the C-terminal cysteine-rich part of the human MUC5AC mucin*. Febs J, 2008. **275**(3): p. 481-9.
113. Maitra, A. and R.H. Hruban, *Pancreatic cancer*. Annu Rev Pathol, 2008. **3**: p. 157-88.
114. Kloppel, G. and N.V. Adsay, *Chronic pancreatitis and the differential diagnosis versus pancreatic cancer*. Arch Pathol Lab Med, 2009. **133**(3): p. 382-7.

115. Horwhat, J.D. and F.G. Gress, *Defining the diagnostic algorithm in pancreatic cancer*. Jop, 2004. **5**(4): p. 289-303.
116. Goonetilleke, K.S. and A.K. Siriwardena, *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): p. 266-70.
117. Barton, J.G., et al., *Predictive and prognostic value of CA 19-9 in resected pancreatic adenocarcinoma*. J Gastrointest Surg, 2009. **13**(11): p. 2050-8.
118. Kalthoff, H., et al., *Characterization of CA 19-9 bearing mucins as physiological exocrine pancreatic secretion products*. Cancer Res, 1986. **46**(7): p. 3605-7.
119. Yue, T., et al., *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): p. 1697-707.
120. Haab, B.B., et al., *Glycosylation Variants of Mucins and CEACAMs as Candidate Biomarkers for the Diagnosis of Pancreatic Cystic Neoplasms*. Annals of Surgery, 2010. **251**(5): p. 937-945.
121. Pines, J.M., *Trends in the rates of radiography use and important diagnoses in emergency department patients with abdominal pain*. Med Care, 2009. **47**(7): p. 782-6.
122. Schweitzer, B., et al., *Multiplexed protein profiling on microarrays by rolling-circle amplification*. Nat Biotechnol, 2002. **20**(4): p. 359-65.
123. Zhou, H., et al., *Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements*. Genome Biol, 2004. **5**(4): p. R28.
124. Schmidt, B.F., et al., *Signal amplification in the detection of single-copy DNA and RNA by enzyme-catalyzed deposition (CARD) of the novel fluorescent reporter substrate Cy3.29-tyramide*. J Histochem Cytochem, 1997. **45**(3): p. 365-73.
125. Rissin, D.M., et al., *Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations*. Nat Biotechnol. **28**(6): p. 595-9.

126. Jacobs, C.B., M.J. Peairs, and B.J. Venton, *Review: Carbon nanotube based electrochemical sensors for biomolecules*. Anal Chim Acta. **662**(2): p. 105-27.
127. Haab, B.B., *Antibody-lectin sandwich arrays for biomarker and glycobiology studies*. Expert Rev Proteomics, 2010. **7**(1): p. 9-11.
128. Ringel, J. and M. Lohr, *The MUC gene family: Their role in diagnosis and early detection of pancreatic cancer*. Mol Cancer, 2003. **2**(1): p. 9.
129. Adsay, N.V., *Role of MUC genes and mucins in pancreatic neoplasia*. Am J Gastroenterol, 2006. **101**(10): p. 2330-2.
130. Zeng, Z., et al., *The development of an integrated platform to identify breast cancer glycoproteome changes in human serum*. J Chromatogr A, 2009: p. In press.
131. Perou, C.M., et al., *Molecular portraits of human breast tumours*. Nature, 2000. **406**: p. 747-752.
132. Iacobuzio-Donahue, C.A., et al., *Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies*. Cancer Res, 2003. **63**(24): p. 8614-22.
133. Logsdon, C.D., et al., *Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer*. Cancer Res, 2003. **63**(10): p. 2649-57.
134. Pepe, M.S., et al., *Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design*. J Natl Cancer Inst, 2008. **100**(20): p. 1432-8.
135. Pleskow, D.K., et al., *Evaluation of a serologic marker, CA19-9, in the diagnosis of pancreatic cancer*. Ann Intern Med, 1989. **110**(9): p. 704-9.
136. Ko, A.H., et al., *Serum CA19-9 response as a surrogate for clinical outcome in patients receiving fixed-dose rate gemcitabine for advanced pancreatic cancer*. Br J Cancer, 2005. **93**(2): p. 195-9.

137. Ritts, R.E., Jr., et al., *Comparison of preoperative serum CA19-9 levels with results of diagnostic imaging modalities in patients undergoing laparotomy for suspected pancreatic or gallbladder disease*. Pancreas, 1994. **9**(6): p. 707-16.
138. Yue, T. and B.B. Haab, *Microarrays in glycoproteomics research*. Clin Lab Med, 2009. **29**(1): p. 15-29.
139. Wu, Y.M., et al., *Mucin glycosylation is altered by pro-inflammatory signaling in pancreatic-cancer cells*. J Proteome Res, 2009. **8**(4): p. 1876-86.
140. Mahley, R.W., et al., *Plasma lipoproteins: apolipoprotein structure and function*. J Lipid Res, 1984. **25**(12): p. 1277-94.
141. Akagi, J., et al., *CA19-9 epitope a possible marker for MUC-1/Y protein*. Int J Oncol, 2001. **18**(5): p. 1085-91.
142. Young, S.G., *Recent progress in understanding apolipoprotein B*. Circulation, 1990. **82**(5): p. 1574-94.
143. Kane, J.P., D.A. Hardman, and H.E. Paulus, *Heterogeneity of apolipoprotein B: isolation of a new species from human chylomicrons*. Proc Natl Acad Sci U S A, 1980. **77**(5): p. 2465-9.
144. Haab, B.B., et al., *Glycosylation variants of mucins and CEACAMs as candidate biomarkers for the diagnosis of pancreatic cystic neoplasms*. Ann Surg, 2010. **251**(5): p. 937-45.