Diagnostic and Prognostic Capacity of Serum Glycan Nodes in Different Types of Cancer

by

Shadi Roshdiferdosi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2018 by the
Graduate Supervisory Committee:

Chad R. Borges, Chair
Neal Woodbury
Mark Hayes

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Glycans are monosaccharide-based heteropolymers that are found covalently attached to many different proteins and lipids and are ubiquitously displayed on the exterior surfaces of cells. Serum glycan composition and structure are well known to be altered in many different types of cancer. In fact, glycans represent a promising but only marginally accessed source of cancer markers. The approach used in this dissertation, which is referred to as “glycan node analysis”, is a molecularly bottom-up approach to plasma/serum (P/S) glycomics based on glycan linkage analysis that captures features such as α2-6 sialylation, β1-6 branching, and core fucosylation as single analytical signals.

The diagnostic utility of this approach as applied to lung cancer patients across all stages as well as prostate, serous ovarian, and pancreatic cancer patients compared to certifiably healthy individuals, nominally healthy individuals and/or risk-matched controls is reported. Markers for terminal fucosylation, α2-6 sialylation, β1-4 branching, β1-6 branching and outer-arm fucosylation were most able to differentiate cases from controls. These markers behaved in a stage-dependent manner in lung cancer as well as other types of cancer. Using a Cox proportional hazards regression model, the ability of these markers to predict progression and survival in lung cancer patients was assessed. In addition, the potential mechanistic role of aberrant P/S glycans in cancer progression is discussed.

Plasma samples from former bladder cancer patients with currently no evidence of disease (NED), non-muscle invasive bladder cancer (NMIBC), and muscle invasive bladder cancer (MIBC) along with certifiably healthy controls were analyzed. Markers for α2-6 sialylation, β1-4 branching, β1-6 branching, and outer-arm fucosylation were able to separate current and former (NED) cases from controls; but NED, NMIBC, and MIBC

were not distinguished from one another. Markers for α2-6 sialylation and β1-6 branching were able to predict recurrence from the NED state using a Cox proportional hazards regression model adjusted for age, gender, and time from cancer. These two glycan features were found to be correlated to the concentration of C-reactive protein, a known prognostic marker for bladder cancer, further strengthening the link between inflammation and abnormal plasma protein glycosylation.

DEDICATION

To my mom, Nasrin, my role model, my hero, the strongest woman I have ever known.

To the memory of my dad, Hossein, who taught me chemistry before I knew alphabets.

To my family, whose love has no boundaries.

To Ali, who taught me anything is possible.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Chad R. Borges, for his exceptional guidance and support in my graduate research and study. Without his mentorship and counsel this dissertation would not have been possible.

I would also like to thank my committee members, Dr. Neal Woodbury and Dr. Mark Hayes for their guidance over the years.

I thank the School of Molecular Sciences and Biodesign Center for Personalized Diagnostics, The Biodesign Institute at Arizona State University for providing the opportunity to work with a great group of scientists and outstanding facilities. I would like to thank Dr. Ian Gould for his support from the first day we joined this program. In addition, I would like to thank Douglas Rehder who taught me a lot of lab techniques when I first joined my lab group. I am thankful to my lab mates and friends, Yueming Hu, Stephanie Thibert, Erandi Kapuruge, Joshua Jeffs, and Jesus Aguilar; as well as members of Center for Molecular Biomarkers, Olgica Trenchevska and Dobrin Nedelkov, for their valuable ideas and suggestions whenever I faced a problem in my research, and for keeping me motivated even on the hard days of my PhD life.

I am grateful to the National Cancer Institute of the National Institute of Health and Flinn Foundation for their financial contribution in my research.

Last but not least, my heartfelt thanks to my family for their unconditional love and support, for always being there for me even though they were thousands of miles away; to Ali, my husband, for his endless love and encouragement throughout my graduate studies; to my sister, Shayesteh, for being by my side during my PhD career and making it easier to be far from my family.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATION

AFP	α-fetoprotein
AUC	Area under the curve
bFGF	Basic fibroblast growth factor
CA125	Carcinoma antigen 125
CE	Capillary electrophoresis
CRP	C-reactive protein
CT	Computed tomography
DMSO	Dimethyl sulfoxide
EPA	Epithelial adhesion
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
FA2	Fucosylated, agalactosyl biantennary
GAG	Glycosaminoglycan
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GBS	Group B Streptococcus
GC	Gas chromatography
GlcNAc	N-Acetylglucosamine
GPI	Glycosylphosphatidylinositol
GT	Glycosyltransferase
HCC	Hepatocellular carcinoma
HexNAc	N-Acetylhexosamine
HILIC	Hydrophilic interaction liquid chromatography
IgG	Immunoglobulin G

IRB	Institutional Review Board
LC	Liquid chromatography
LCO	Lipo-chitooligosaccharide
MCFD2	Multiple Coagulation Factor Deficiency 2
MHC	Major histocompatibility complex
MIBC	Muscle-invasive bladder cancer
MMP	Matrix metalloproteinase
MS	Mass Spectrometry
NED	No evidence of disease
Neu5Ac	N-acetylneuraminic acids
NK	Natural killer
NMIBC	Non-muscle-invasive bladder cancer
P/S	Plasma/serum
PAMP	Pathogen associated molecular pattern
PGC	Porous graphitized carbon
PMAA	Partially methylated alditol acetate
ppGalNAcT	Polypeptide GalNAc transferase
PSA	Prostate-specific antigen
QC	Quality control
ROC	Receiver operating characteristic
SAMP	Self-associated molecular pattern
Sia6LacNAc	α2,6-sialylated lactosamine
SLe^a	Sialyl Lewis a
ST6Gal-I	β-galactoside α2,6-sialyltransferase I
t-Glc	Terminal glucose

TFA	Trifluoroacetic acid
TGF-β1	Transforming growth factor beta 1
UCC	Urothelial cell carcinoma
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine

CHAPTER 1

INTRODUCTION

1.1 The Structure, Biological Function, and Importance of Glycans

Glycans are one of the abundant and diverse natural biopolymers. In biological systems, glycans are attached to proteins and lipids displayed on exterior surface of cells and macromolecules consisting variety of structures. Glycans are produced by glycosyltransfeases in the endoplasmic reticulum (ER) and golgi apparatus organelles. Studies on glycan functions show that enzymes such as glycosyltransferases and glycosidases have important role in the development and physiology of living organisms [1]. Glycans are involved in such diverse important biological processes as cell adhesion, signal transduction, molecular recognition, receptor activation, endocytosis, and molecular trafficking and clearance. Glycobiology is the field of studying and understanding the structure and biological function of glycans. The science of protein expression, modification, interaction and function in biological systems, proteomics, has had a major growth over the years; however, proteins' post-translational modifications which happen in majority of proteins in eukaryotic cells have not been significantly studied until recently [2]. One of the most common and complex post-translational modifications is glycosylation, which has an essential role in cell viability.

1.1.1 Types of Glycans

The three major classes of glycans include N-linked glycans, O-linked glycans, and glycolipids (**Figure 1.1**). N-linked glycans are attached to the nitrogen atom of aspargine residues with an amino acid sequence of Asn-X-Ser/Thr where X can be any amino acid except proline. O-linked glycans are usually attached to the oxygen atom of serine or threonine residues on the protein. A common type of protein O-glycosylation is mucine

type O-glycosylation in which the O-glycan is attached via N-acetylgalactosamine (O-GalNAc). There are also other types of O-glycosylation that are initiated via mannose or N-acetylglucosamine (GlcNAc). Moreover, some forms of glycosylation only occur in specific types of proteins, such as the Notch receptors that are usually modified with different O-glycans including O-fucose (**Figure 1.1**) [3]. Lipid-linked glycans are usually attached to a sphingolipid or a glycerol group that consists of one or two fatty acids. Glycosylphosphatidylinositol (GPI) is a lipid-linked glycan that is usually attached to the C-terminus of a protein. These GPI-anchored proteins can be found on the cell plasma membrane (**Figure 1.1**). There is another class of glycoconjugates called proteoglycans that have one or multiple glycosaminoglycans (GAGs) (**Figure 1.1**). Heparan sulphate, hyaluronan, and chondroitin sulfate are some of the most important GAGs in human body.

In order to analyze the peptides or the glycans attached to glycoproteins, removing the oligosaccharide structures from glycoproteins facilitates the process. There are highly efficient methods, both chemical and enzymatic, to separate oligosaccharides from glycoproteins. Chemical methods usually cause degradation of proteins. For instance, in solid phase permethylation, O-glycans are released from proteins using iodomethane in sodium hydroxide beads but it also results in degradation of the protein [4]. Enzymatic methods are softer compared to chemical methods. For example, to release N-glycans from glycoproteins, PNGase F enzyme is the most efficient. It releases the N-glycans by hydrolyzing the glycan chain from glycoproteins as well as deamination of asparagine to aspartic acid [5]. Enzymatic methods of releasing O-glycans are more challenging. Most often, a series of exoglycosidase digestions or a combination of enzymatic/chemical release must be carried out to release O-glycans from glycoproteins [4]. Song et al. have recently developed a simple approach to release N- and O-glycans

from kilogram amounts of biological samples [6]. In this method, which is termed "Oxidative release of Natural Glycans", biological samples are treated in a controlled condition with sodium hypochlorite (NaClO) found in commercial bleach to selectively release N- and O-Glycans, which then can be tagged specifically for chromatographic separation and structural elucidation [6].

1.1.2 Biological Functions of Glycans

In order to understand the biological functions of glycans a variety of approaches can be taken including chemical or enzymatic removal of the intact glycan structures from glycoproteins, manipulation of glycan processing, and elimination of glycosylation sites genetically. Based on the results of the aforementioned approaches taken by scientists over the years, glycan roles can be classified into four general categories: structural and modulatory roles, exogenous (extrinsic) recognition, endogenous (intrinsic) recognition, and molecular mimicry of host glycans (**Figure 1.2**) [1]. Each of these roles is discussed with examples here.

1.1.2.1 Structural and Modulatory Roles

Structural features of glycans mediate some essential biological effects since glycans have a broad presence with considerable abundance in intra- and extracellular areas and body fluids. For instance, the toughness and strength of plants and cell walls of fungi come from some glycan polymers such as cellulose and chitin, which are β-linked homopolymers of glucose and N-acetylglucosamine, respectively [7], [8]. In another example, glycocalyx, a heterogeneous group of glycoprotein covering all eukaryotic cells, has a major role as a physical barrier [9], [10].

Furthermore, glycans are able to alter the function of proteins and lipids that are attached to them. Glycoprotein folding is one of the mechanisms in which glycans have an important role. More than 50% of the proteins synthesized via secretory pathway are glycosylated. Glycosylation of these proteins facilitates their folding. Most of the ER-synthetized proteins are glycosylated by N-glycans. Considering their large hydrophilic structures, these N-glycans should be involved in proper folding of the synthetized polypeptides. In 1977, Leavitt et al. examined the effect of N-linked glycans on folding of these proteins [11]. Unlike the glycosylated forms, the nonglycosylated glycoproteins of vesicular stomatitis virus and sindbis virus could not be detected on the outside of the plasma membrane. The inhibition of N-linked glycosylation of these glycoproteins using tunicanycin resulted in improper folding of such proteins, leading to their inability to migrate to the cell surface [11].

Studies indicate that not only can the signaling characteristics of protein receptors be affected by glycoproteins and glycolipids that are in the same cell membrane, but also they can change by glycans attached to these proteins and lipids. In 1991, Rapraeger et al. found out that by preventing heparan sulfate at the cell surface to bind to basic fibroblast growth factor (bFGF), the binding of FGF to its receptors with tyrosine kinase activity significantly decreases [12]. As another example, a study in 2005 by Wang et al. showed alteration in core fucosylation of N-glycans of TGF $-\beta 1$ receptors can lead to emphysema by causing over-expression of matrix metalloproteinases (MMPs) [13].

Another function of glycans is protection from immune recognition. The foreign proteins can be recognized by B cell receptors. Also, they can be coupled with major histocompatibility receptors following recognition by T cell receptors. Peptides containing small glycans can be recognized by glycopeptide-specific cytotoxic T cells

[14]; however, peptides with larger glycans can usually escape the major histocompatibility coupling and T cell receptor recognition. For instance, influenza hemagglutinin that is heavily glycosylated uses this strategy to evade immune system [15].

The N-glycans attached to glycoproteins on cell surfaces can have different number of branches. Not only glycan branching can change the structure of the protein that it is attached to, but also it can change various biological functions. As Demetriou et al. reported in 2001, N-glycan branching increases activation of T cell [16]. In addition, β1-6 branching is directly involved in malignant transformation. It is shown that introducing increased β1-6 branching to mouse models with nonmetastatic breast cancer significantly increased their metastatic potential [17]. Another type of N-glycan branching that is called bisecting GlcNAc, is involved in inhibition of growth factor signaling, which leads to reducing the progression of mammary tumor [18]. E-cadherin, showing this type of abnormal glycosylation, is classified as a tumor suppressor and it is reported as a potential diagnostic biomarker in breast cancer [19]. N-glycan branching occurs by addition of β-linked GlcNAc residues that are provided by UDP-GlcNAc. The concentration of UDP-GlcNAc is related to the distribution of glycoproteins at the cell surface, cellular metabolism, and some diseases since glucosamine and GlcNAc have important metabolic effects in most cells [20]. Lau et al. reported that the number of N-glycans and degree of branching on a glycoprotein is associated with cellular transition between growth and cell cycle arrest and differentiation [21].

1.1.2.2 Exogenous Recognition

Over the time, a lot of pathogens and symbionts have evolved glycan-binding proteins to recognize specific glycans on the complex cell surface of the host species. These

interactions that usually occur through glycan binding proteins can cause symbiosis or disease. For instance, *Helicobacter pylori* that has an important role in gastric ulcer and cancer, recognizes the gastric sialoglycans adhering to the gastric epithelial lining. This bacterial adherence can be mediated by fucosylated Lewis b histo-blood group antigen [22]. One of the well-characterized bacterial adhesins is FimH, which mediates the adhesion of uropathogenic *E coli* to urothelial cells by binding to D-mannose. Uropathogenic *E coli* is accountable for urinary tract infections that are mostly treated with antibiotics. Since antibiotics develop high risk of resistance, Jiang et al. proposed a new class of antimicrobials, which target FimH. Based on their experiments, indolinyphenyl mannoside was the most efficient antimicrobial targeting FimH, reducing the colony-forming units in the bladder of mouse models significantly compared to the common treatment with ciprofloxacin antibiotic [23].

In addition to bacterial adhesins, parasite and fungal adhesins are responsible for some serious diseases. For example, *plasmodium falciparum* malaria parasites infect the red blood cells by recognizing N-acetylneuraminic acids (Neu5Ac) attached to glycophorins on the target red blood cells [24]. As fungal adhesins, epithelial adhesion (EPA) family that are encoded mostly by the *Candida glabrata* genome, are accountable for mediating attachment to host cells. Epa1, Epa6, and Epa7 are adhesins of *Candida glabrata* that recognize ligands containing a terminal galactose residue [25].

Glycans can mediate the adjustment of host immune responses to parasites or symbionts. For instance, polysaccharide A that comes from major mammalian gut microbiome members helps cellular and physical maturation of immune system by improving T cell deficiencies and controlling lymphoid organogenesis [26]. Similarly, glycans expressed by parasitic helminthes interact with host glycan binding proteins to

adjust innate and adaptive immune responses in order to promote their survival, a concept known as “glycan gimmickry” [27].

Glycans can facilitate antigen recognition, uptake, and processing. As an example, foreign proteins with significant amount of terminal mannose and N-acetyl glucosamine (GlcNAc) residues attached to them can activate phagocytosis through C-type lectin receptors expressed on antigen-presenting cells resulting in activation of immune responses [28]. As another form of non-self recognition, CD1 molecules that present foreign glycolipids can be recognized by invariant T cell antigen receptor expressed by natural killer T lymphocytes [29].

1.1.2.3 Endogenous Recognition

Glycans function as specific ligands not only for cell-microbe interactions but also for cell-cell interaction within the same species. The first glycan receptors discovered with intrinsic recognition were those containing mannose-6-phosphate, which are recognized by P-type lectins leading to mediation of lysosomal enzyme trafficking to lysosomes [30]. So far, various glycan-binding proteins have been discovered that have intrinsic functions mediating different biological processes.

Aside from their role in nascent glycoprotein folding mentioned earlier, glycans are involved in ER associated degradation (ERAD). For example, the lipid-linked oligosaccharide donor for N-glycosylation that has the same structure ($Glc_3Man_9GlcNAc_2$-P-P-dolichol) [31] in nearly all eukaryotes plays an important role during glycoprotein folding in ER. In 1983, Parodi et al. found a glucosylation/deglucosylation cycle in this process in which the third glucose residue is regularly detached and then attached again [32]. Later in 1994, it was discovered that calnexin and calreticulin, ER chaperones, recognize monoglucosydated oligosaccharides

on glycoproteins [33]. After removal of the last glycose residue, the oligomannose N-glycans can be recognized by mannose 6-phosphate receptors of some chaperone proteins [34]. Ultimately, a series of N-glycan dependent quality controls determine whether a glycoprotein will continue in the secretory pathway or will be sent for ERAD.

Similar to mannose 6-phosphate recognition by P-type lectins for lysosomal enzyme trafficking, there are other molecules like lectin in the secretory pathway that regulate intracellular glycoprotein trafficking. For instance, ERGIC-53 that is a mannose-binding and calcium-dependent human homolog of leguminous lectins localized to the ER-Golgi intermediate compartment, when forming a complex with MCFD2, builds a specific cargo receptor for ER to Golgi transportation of selected glycoproteins such as factor V and factor VIII clotting proteins [35].

Glycans are also involved in triggering endocytosis. Different known lectins perform endocytosis in macrophages and antigen-presenting cells. In addition to T cell activation, this process is important for damaged cells and glycoprotein clearance. For example, in case of cancer and inflammation heterogeneously O-glycosylated mucins aberrantly enter the bloodstream; however, the circulating mucins that are used as clinical diagnostic markers are only the clearance-resistant subset of these molecules and most of the secretory mucins entering the circulation are cleared by multiple glycan recognizing hepatic clearance receptors [36].

The role of glycans in intercellular signaling is essential. Bacterial Nod factors that are lipo-chitooligosaccharides (LCO) are well known for communicating signals produced by rhizobial bacteria and recognized by lysine-motif receptors in the roots of legumes. This process activates the common symbiotic pathway, which eventually results in production of nodules on the root and nitrogen fixation. Nod factors also participate in recognition of plant pathogenic microorganisms [37].

Intercellular adhesion is among important glycan functions. Selectins are transmembrane glycoproteins that bind to focusylated and sialylated glycoproteins mediating leukocytes adhesion to the vascular endothelium. Since selectin interactions have an important role in cancer and inflammation, targeting them for treatment of such pathological conditions can be useful. For example, oral use of fucose proved helpful in correction of leukocyte adhesion deficiency-II by increasing the expression of fucosylated selectin ligands on the surface of neutrophils [38]. In another example, the pan-selectin inhibitor GMI-1070 reduces selectin mediated cell adhesion and abrogates sickle cell crisis in animal models [39].

1.1.2.4 Molecular Mimicry of Host Glycans

Glycans can serve as pathogen associated molecular patterns (PAMPs), which can be detected by pattern recognition receptors of innate immune cells such as NOD-like receptors and C-type lectins. Glycans can also inhibit immune response by acting as self-associated molecular patterns (SAMPs) that are recognized by intrinsic inhibitory receptors. The evolution of microorganisms has resulted in phenomenal molecular mimicry of host glycans. For instance, the structure of Group B *Streptococcus* (GBS) polysaccharides include Neu5Acα2-3Galβ1-4GlcNAcβ1- unit, which is in common with the structure of host glycans. It has been shown that Siglec-9 interacts with sialic acids on GBS the same way it does with host sialic acids, inhibiting neutrophil responses [40].

Sialoglycan mimicry by microorganisms can be done in different ways. Some of them like *Psseudomonas aeruginosa* adsorb sialoglycoproteins [41] and some others such as *Haemophilus influenzae* acquire the free sialic acids in their environment [42]. Also, there is another group of microorganisms like *Trypanosoma cruzi* that use trans-

sialidases to transfer host sialic acids to the terminal β-galactopyranosyl residues of mucin-like molecules on the cell surface of the parasite [43].

1.2 Alteration of Glycans in Cancer

Nearly all tumor cells display aberrant glycan structures. Since abnormal glycan modifications are unique for each protein, glycosylation site, and cell, the molecular and functional diversity increases within cell population. There are different factors in glycosylation process that affect formation of the final glycan structure in a specific cell or tissue type. The two major concepts for tumor-associated alterations of glycan structures are incomplete synthesis and neo-synthesis, which were formulated by Hakomori et al. in 1983. The incomplete synthesis process usually happens in early stages of cancer as a result of the destruction of normal synthesis of glycan structures expressed in normal epithelial cells, leading to formation of truncated structures. For example, abnormal synthesis of sialyl-Tn antigen, which occurs due to the overexpression of ST6GalNAc I, changes the glycosylation pattern of tumor cells in breast cancer, resulting in decreasing their adhesion and increasing their migration [44]. Neo-synthesis that most often occurs in late stage cancers is a consequence of induction of transcription of certain important glycogenes. For instance, expression of sialyl Lewis a (SLe^a) and SLe^x is induced in advanced stages of several cancers by hypoxia induced transcription of the genes involved in the expression of these carbohydrate determinants [45].

One of the important upstream causes of altered glycans expression is under- or overexpression of glycosyltransferases (GTs), which is a result of transcriptional dysregulation [46], chaperone dysfunction [47], and/or altered glycosidase activity [48]. In fact, dysregulation of glycosyltrasferase activity is a hallmark of almost every known

type of cancer [49]. In addition, changes in the tertiary structure of the peptide backbone and/or the attached glycan chain can lead to altered expression of glycans. Besides, since individual glycosyltrasferase enzymes show strict donor, acceptor, and linkage specificity, the acceptor substrate, availability and abundance of the sugar-nucleotide donors, and expression and localization of proper GTs in the Golgi apparatus can contribute to changing the glycan expression [50].

Glycosylation changes that are related to cancer can be divided into five categories: sialylation, fucosylation, N-glycan branching and bisecting GlcNAc, and O-glycan truncation.

Sialylation plays an essential role in cellular recognition, cell signaling and cell adhesion. It has been shown that in cancer, abnormal GT expression leads to an increase in α2,6- and α2,3-linked sialylation [51]. For example, the expression of β-galactoside α2,6-sialyltransferase I (ST6Gal-I) that synthesizes α2,6-sialylated lactosamine (Sia6LacNAc) is altered in different cancers such as colon, stomach, and ovarian cancer and has been shown to be a predictive marker of poor prognosis in colon cancer [52], [53]. Another important sialylated antigen is SLe^x, which is overexpressed in several malignant cancers and its expression level can be predictive of patients' survival [54]. Overexpressed SLe^x promotes the development of metastasis by interacting with selectins in cancer, forming bulks of cancer cells and facilitating their adhesion to endothelia [55].

Another glycosylation factor associated with cancer is fucosylation. There are two types of fucosylations: terminal fucosylation and core fucosylation. In the biosynthesis of SLe antigens, the terminal steps consist of α1,3- or α1,4-focusylation of SLe^a or SLe^x that are α2,3-sialylated [56]. It has been reported that the increased expression of SLe^x in breast tumors is dependent mainly on the activity of Fuc-TVI, a fucosyltransferase

enzyme encoded by FUT6 [57]. As another example, Fuc-TVII regulates the expression of SLe^x in adult T cell leukaemia cells. The human T-lymphotropic virus I retrovirus encodes TAX, a transcriptional activator protein, which modulates the FUT7, the gene that encodes Fuc-TVII [58]. In core fucosylation, α1,6-fucose is added to the GlcNAc residue at the N-glycans core by Fuc-TVIII, a fucosyltransferase encoded by FUT8. In several cancers including lung cancer, enhanced expression of FUT8 and core fucosylation is a major feature [59]. Core fucosylation of α-fetoprotein in hepatocellular carcinoma (HCC) facilitates the early diagnosis of this disease. In fact, it is an approved biomarker for HCC that can distinguish it from chronic hepatitis and liver cirrhosis [60].

Branching and bisecting GlcNAc N-glycans are also among common feature of tumors. During malignant transformation the activity of GnT-V increases. As a result, β1,6-branched N-linked glycans are overexpressed in cancer cells. It has been shown that the gene encoding GnT-V, MGAT5, is regulated by the RAS-RAF-MAPK signaling pathway, which is activated in cancer [17]. For instance, an immortalized lung epithelial cell line transfected with a GnT-V expressing vector exhibited characteristics of cellular transformation such as loss of contact inhibition, increased cell motility and tumor formation in nude mice [61]. Moreover, downregulation of GnT-V in cancer cell lines leads to suppression of tumor progression and metastasis. Mammary tumor growth and metastases induced by a viral oncogene was significantly less in Mgat5 deficient mice than in transgenic littermates expressing Mgat5 [62]. GnT-III, which is encoded by MGAT3 and is responsible for adding bisecting GlcNAc N-glycans in a β1,4-linkage, has a tumor suppressive role as opposed to GnT-V. For example, transfection of MGAT3 into highly metastatic murine melanoma cells resulted in a significant suppression of lung metastasis in syngeneic and nude mice. Also, due to GnT-III and GnT-V competition for substrate β1,6-branching was significantly decreased [63].

Overexpression of truncated O-glycans is another characteristic of tumors. The mucin-type O-glycans, which are initiated by polypeptide GalNAc transferases (ppGalNAcTs) are usually found in most transmembrane and secreted glycoproteins [64]. The expression of ppGalNAcTs is usually altered in cancer leading to abnormal O-glycosylation [65]. Furthermore, competition of enzymes for the same substrate can lead to overexpression of truncated glycans, exposing protein epitopes that would be hidden in normal conditions. For instance, the relative enzymatic activities of C2GnT1 and ST3Gal-I indicate the dominant O-glycan structure and control the expression of a tumor-associated epitope, SM3, on MUC1 [66]. It has been shown that truncated O-glycan structures Tn and Sialyl-Tn are associated with oncogenic features, such as enhanced tumor growth, decreased cell adhesion and invasion [67]. Since Sialyl-Tn can be detected in most carcinomas as opposed to healthy tissues, it has been considered as a potential prognostic marker and a target of cancer vaccines [68].

1.3 Glycans in Cancer Diagnosis

Cancer is one of the leading causes of death in the world. There is an essential need for new approaches toward risk prediction, early diagnosis, and treatment of cancer. Recently, glycans have proven to be a promising source for the development of new non-invasive biomarkers.

Currently, some of the clinically approved biomarkers that are being used for cancer diagnosis and prognosis are glycoproteins [69]. For example, prostate-specific antigen (PSA) is a biomarker widely used for prostate cancer diagnosis [70]. Other instances of glycoprotein biomarkers include: carcinoma antigen 125 (CA125) for ovarian cancer [71], SLe^a and CA19-9 for colon, gastric, and pancreatic cancers [69], [72], and aberrantly glycosylated MUC1, also known as CA15-3, for breast cancer [73]. Due to

their relatively low specificity, all of the aforementioned serological biomarkers have restricted application. This issue has encouraged a search for new biomarkers with higher specificity and sensitivity for early diagnosis of cancer based on detection and quantification of specific glycan structures attached to a particular protein. For instance, α-fetoprotein (AFP) that is biomarker for detection of liver diseases [60] is not suitable for distinguishing between HCC and noncancerous liver diseases in serum levels; however, a glycosylated form of AFP, the AFP-L3 fraction, is significantly fucosylated in HCC compared to other liver diseases, making it an FDA approved marker for early diagnosis of HCC [74].

Developing new technologies and novel glycan analysis techniques has facilitated the discovery of many putative glycan-based markers. Because of its sensitivity, possibility for high throughput, and tolerance for complex biological samples, mass spectrometry is one of the main methods of glycosylation analysis and has increasingly been used in cancer biomarkers discovery [75]. In order to detect and measure potential cancer markers, biological samples generally need a pre-separation method including liquid chromatography (LC), gas chromatography (GC), ion mobility, and capillary electrophoresis (CE) before the MS analysis. The common techniques to study modifications of glycan structures during development and progression of different types of cancer as described in recent reviews are including LC-ESI-MS and MALDI-MS [76], [77]. In a study by Kyselova et al. in 2008, it has been shown that N-glycomic profiling of serum components using mass spectrometry is a sensitive and informative approach for tracking the progression of cancer. Performing a specific MALDI-MS glycomic profile analysis on permethylated glycans in sera from breast cancer patients in different stages of cancer alongside healthy individuals, they found significant increase in sialylation and fucosylation of glycan structures associated with cancer progression [78]. In another

study, a chip-based reversed-phase LC-ESI-MS method was applied to reduced and permethylated N-glycans extracted from breast cancer samples. In line with other breast cancer studies, it was demonstrated that branching and sialylation of glycans increase with progression of cancer [79].

In 2011, Arnold et al. applied a high-throughput HILIC technology on serum samples donated by 100 lung cancer patients (20 samples in each stage I, II, IIIA, IIIB, and IV) as well as 84 age-matched controls. Significant increases in levels of SLe^x, mono-antennary, and trisialylated glycans were detected in lung cancer samples compared to controls. The combination of these significantly increased glycans resulted in higher sensitivity and specificity. In addition, the N-glycan profile of whole serum glycome was compared to that of isolated haptoglobin and they both had similar glycosylation alteration patterns [80].

The levels of glycosylation of glycoproteins of bladder cancer cell lines were examined by MALDI-TOF/TOF-MS. As a result, SLe^x, terminal GalNAc and Gal, and high mannose-type N-glycans showed significant increases in bladder cancer cells and tissues. Furthermore, the expression of core-fucosylated N-glycans were higher in bladder cancer cells but that of terminally fucosylated N-glycans were lower [81].

Glycosylation changes were assessed in ovarian cancer using HPLC glycomic profiling with fluorescence detection and MS. The levels of core fucosylated, agalactosyl biantennary glycans (FA2) and SLe^x were significantly increased in ovarian cancer patients. In this study, they showed that the level of SLe^x was elevated in the acute-phase proteins, haptoglobin, α1-acid glycoprotein, and α1-antichymotrypsin compared to healthy controls. In addition, the immunoglobulin G (IgG) contained significantly higher level of FA2 [82]. Other studies [83], [84] are in agreement with the results of this study, which suggest the independence of abnormal glycosylation from the analytical method.

In a pancreatic cancer study, after purifying and analyzing haptoglobin using LC-ESI-MS and MALDI-TOF, it was indicated that α1,3-, α1,4-, and α1,6- fucosylation of this protein was significantly high among pancreatic cancer patients [85]. In another study in 2011, Lin et al. identified a bifucosylated triantennary structure for the first time in pancreatic cancer samples by extracting haptoglobin from serum, permethylating the desialylated N-glycans, and analyzing them by MALDI-QIT-TOF-MS. Comparing the results of this analysis applied on sera from 16 pancreatic cancer patients and 15 individuals with benign conditions, they found higher level of both core and antennary fucosylation in pancreatic cancer samples [86]. However, the sample size in this study is small and the results have not been validated yet.

N-glycans derived from blood serum of 24 prostate cancer patients and 10 healthy individuals were permethylated and analyzed by MALDI-MS. A total of 12 glycan structures were found to be significantly different between cases and controls. Half of these glycans were fucosylated. The levels of high mannose and complex biantennary structures were lower in cancer patients but fucosylated complex biantennary and complex tetraantennary structures were increased in prostate cancer compared to controls [87].

The development of new technologies and methods of separation and mass spectrometry has led to many studies conducted by different groups to demonstrate the correlation of abnormal production of glycans to cancer development and progression. All the methods described here analyzed the intact structure of glycans. We recently developed a molecularly bottom-up approach to plasma/serum (P/S) glycomics based on glycan linkage analysis that captures features such as α2-6 sialylation, β1-6 branching, and core fucosylation as single analytical signals [88]. In Chapter 2 of this dissertation, validation of robustness of this method is discussed. In chapters 3 and 4, this method

was applied to over 1000 P/S samples from patients with different types of cancer across all stages matched with certifiably healthy, nominally healthy and/or risk-matched controls to evaluate the diagnostic and prognostic capacity of glycan “nodes” in different types of cancer.

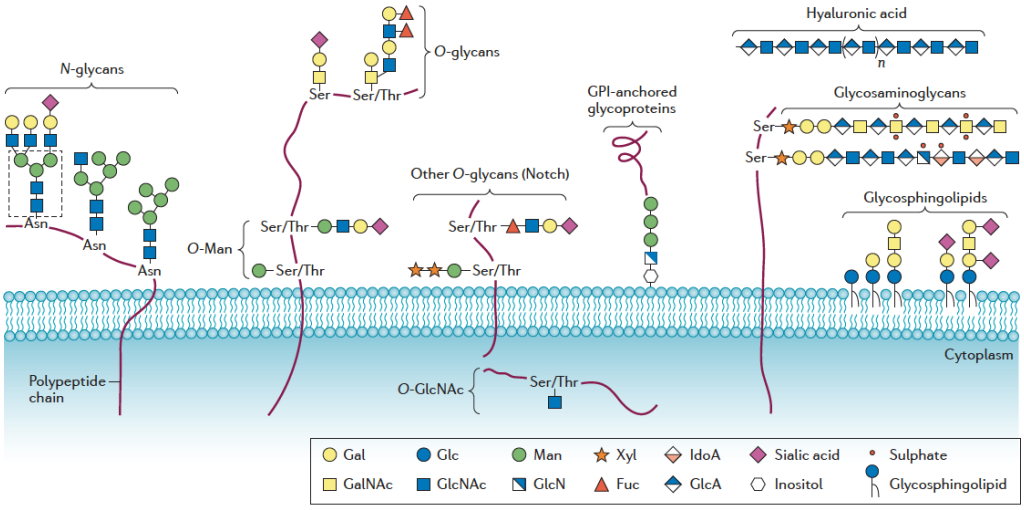

Figure 1.1. Common types of glycans in mammals. Glycans are found covalently attached to many different proteins and lipids and are ubiquitously displayed on the exterior surfaces of cells. N-glycans that are attached to Asp residue of proteins, have a common core that is shown in a dotted line box in the figure. O-glycans that are attached to Ser/Thr residue of proteins can have various cores. Glycosaminoglycans are unbranched co-polymers consisting of repeating acidic disaccharide units that are usually attached to proteoglycans. Glycosphingolipids are a subtype of glycolipids that are found on the cell membrane. Glycosylphosphatidylinositol (GPI)-anchored proteins are glycoproteins attached to a phosphatidylinositol found on the cell plasma membrane. Adopted with permission from Pinho et al. [89].

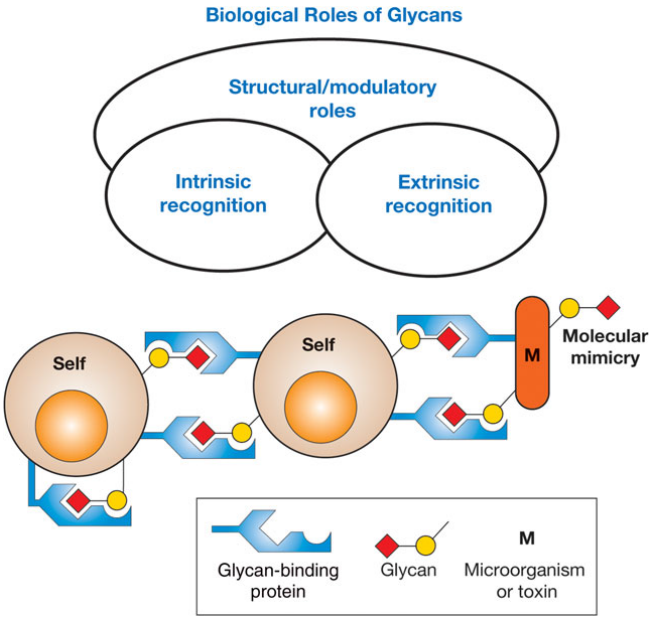

Figure 1.2. General categories of the biological functions of glycans. Endogenous recognition is shown by on the left side of the central "self" cell. The right side binding demonstrates exogenous recognition. Molecular mimicry of host glycans makes the roles of glycans more complicated. In some cases some overlap can be found between different categories. Adapted with permission from Varki [1].

CHAPTER 2

GLYCAN NODE ANALYSIS: VALIDATION OF ASSAY ROBUSTNESS [1]

2.1 Introduction

Serum glycan composition and structure are well known to be altered in many different types of cancer [77], [90]–[92]. In fact, for over a decade now, global blood plasma/serum (P/S) glycomics has held out the promise of new, non-invasive cancer markers derived from a small volume of this easily accessible biofluid [93], [94]. Modern analytical methods for quantifying the relative abundance of different glycans in P/S vary widely [95]–[97], ranging from multiplexed capillary gel electrophoresis with laser-induced fluorescence (a DNA sequencer-adapted method) [98], [99] to hydrophilic interaction liquid chromatography (HILIC) [100] or porous graphitized carbon (PGC) [101], [102] chromatography interfaced with electrospray ionization-based mass spectrometers or as a means of prefractionation prior to analysis by MALDI-MS [103]—for which glycans are generally permethylated prior to analysis [97], [104].

Nearly all approaches employed in P/S glycomics focus on the analysis of intact glycans—most commonly N-linked glycans (generally to the exclusion of O-linked glycans and glycolipids). Quite commonly, accounts of such studies that are focused on cancer conclude by taking a wide-angle view of all intact glycans that were altered in cancer relative to a healthy or benign disease state and reporting unique glycan features such as core fucosylation, bisecting N-acetylglucosamine (GlcNAc), and α2-6 sialylation that were found increased or decreased in cancer [94]. Often these features are then directly connected to the activity of specific glycosyltransferases [105]. In 2013, Borges et al [88] developed a molecularly bottom-up approach to serum glycomics which, following permethylation of an unfractionated P/S sample, employs the principles of

[1] This chapter is partially reprinted with permission from S. Ferdosi *et al.*, "Stage Dependence, Cell-Origin Independence, and Prognostic Capacity of Serum Glycan Fucosylation, β1–4 Branching, β1–6 Branching, and α2–6 Sialylation in Cancer," *J. Proteome Res.*, p. acs.jproteome.7b00672, 2018. Copyright 2018, American Chemical Society

glycan linkage analysis to break down all P/S glycans into monosaccharides in a way that maintains information about which hydroxyl groups of each monosaccharide were connected to other carbohydrate residues in the original glycan polymer [88], [106], [107] (**Figures 2.1 and 2.2**). This mode of P/S sample preparation results in a collection of roughly two dozen partially methylated alditol acetates (PMAAs), each of which represents a unique glycan "node" from the original glycan polymers and can readily be quantified by GC-MS. Uniquely, several PMAAs such as those arising from 2,6-linked mannose, 4,6-linked GlcNAc, and 3,4,6-linked mannose correspond to unique glycan features such as β1-6 branching, core fucosylation, and bisecting GlcNAc, respectively, and capture these unique features as single analytical signals rather than allowing the signal from that feature to be spread across all intact glycans that bear the unique feature (**Figure 2.1**). Similarly, many of the PMAAs serve as excellent surrogates for the activities of the glycosyltransferases (GTs) that produced them—as only one or two known human GTs are capable of producing that particular glycan monosaccharide linkage pattern [88]. In addition, this unique approach to P/S glycomics simultaneously captures information from all major classes of P/S glycans, including N-, O-, and lipid-linked glycans [88].

Initial analytical validation of this methodology was done previously [88]. In order to further validate the robustness of this assay, the glycan "node" analysis procedure was applied to matched sets of P/S samples consisting of serum samples and plasmas with different types of anticoagulants to determine whether subtle differences in sample matrix (i.e., different anticoagulants and serum) have an impact on the glycan node composition of the samples. In addition, in order to test the stability of glycans in various sample handling conditions, glycan nodes were analyzed in different types of samples that were treated in different conditions over a year and their matched controls.

2.2 Materials and Methods

2.2.1 Materials

Heavy, stable-isotope-labeled D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-D7, 97%–98%) was purchased from Cambridge Isotope Laboratories. N-acetyl-D-[UL-$^{13}C_6$]glucosamine was obtained from Omicron Biochemicals, Inc. Methanol was purchased from Honeywell Burdick & Jackson. Acetone was obtained from Avantor Performance Materials. Acetonitrile and dichloromethane were acquired from Fisher Scientific. Chloroform, sodium hydroxide beads (20–40 mesh) DMSO, iodomethane (99%, Cat. No. I8507), trifluoroacetic acid (TFA), ammonium hydroxide, sodium borohydride, and acetic anhydride were obtained from Sigma-Aldrich. Pierce spin columns (0.9 mL volume) including plugs were purchased from ThermoFisher Scientific (Waltham, MA, Cat. No. 69705). GC-MS autosampler vials and Teflon-lined pierceable caps were acquired from ThermoFisher Scientific. GC consumables were purchased from Agilent; MS consumables were obtained from Waters.

2.2.2 Plasma and Serum Samples

The first matched set of P/S samples was including samples from 21 healthy donors. This set consisted of four different types of plasma and a serum sample from each donor. The difference between these four types of plasma was based on the different anticoagulants, which were K_2EDTA, K_3EDTA, Sodium EDTA, and 3.8% Sodium Citrate. The second matched set of P/S samples was including 6 matched-collection aliquots of serum, K_2EDTA plasma, and heparin plasma from a single donor. In each matched set, samples were analyzed and compared to each other to verify the consistency of glycan nodes between the aforementioned types of samples.

In order to test the stability of glycans in various sample handling conditions, 7 different types of samples were treated in different conditions over a year. By the end of 360 day time point, glycan nodes were analyzed in these treated samples and their matched controls that were stored at -80 °C the whole time. The samples were including 2 male and 2 female Na_2EDTA plasmas that were matched to serum, K_2EDTA and Heparin plasma from a separate male donor. The different mistreating conditions were including 10 days at -20 °C, 90 days at -20 °C, 360 days at -20 °C, 2 days at 4 °C, 90 days at 4 °C, and 1 day at 25 °C.

All specimens were stored at -80 °C prior to shipment to ASU and maintained at -80 °C at ASU prior to analysis (except for the ones that were intentionally mistreated for the stability test). All specimens were analyzed blind and in random order. An aliquot of plasma from the same individual donor was analyzed in every batch as a quality control (QC) specimen to ensure batch-to-batch consistency.

2.2.3 Experimental Procedures: Glycan Node Analysis

The global glycan methylation analysis procedure consisted of four main steps; permethylation, trifluroacetic acid (TFA) hydrolysis, reduction of sugar aldehydes, acetylation of nascent hydroxyl groups and final cleanup [88], [106]. Each step is described in detail below.

Permethylation, Non-reductive Release, and Purification of Glycans: A total of 9 µL of P/S was added into a 1.5 mL eppendorf tube followed by 1 µL of a 10 mM solution of heavy-labeled D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-D7, 97%–98%), and N-acetyl-D-[UL-$^{13}C_6$]glucosamine which served as internal standards for relative quantification. Then, 270 µL of dimethylsulfoxide (DMSO) was added to the biological sample and mixed to dissolve completely. Once the sample was fully dissolved, 105 µL of

iodomethane was added to the mixture. This solution was then added to a plugged 1 mL spin column, which contained ~ 0.7 g of sodium hydroxide beads. The NaOH beads had been preconditioned with acetonitrile, and rinsed with DMSO twice before the sample was added. Then, the NaOH column was stirred occasionally for 11 min. When finished, samples were unplugged and spun for 15 s at 5,000 rpm (2,400g) in a microcentrifuge to extract the glycan-containing solution. To wash off all the permethylated glycan, 300 µL of acetonitrile was added to the spin column and then centrifuged for 30 s at 10,000 rpm (9,600g). Then, samples from the first spin-through were placed in a silanized 13 × 100 mm glass test tube containing approximately 3.5 mL of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer and mixed well. Next, the second spin-through was pooled with the rest of the sample, avoiding the white residue at the bottom of the spin column. The test tube was capped and shaken thoroughly after adding 1.2 mL of chloroform to the sample. Liquid/liquid extraction was performed 3 times, saving the chloroform layer. The chloroform layer was then extracted with a silanized pipette, transferred to a silanized glass test tube, and dried under nitrogen at heater-block temperature setting of 74 °C.

TFA hydrolysis: A total of 325 µL of 2 M TFA was added to each sample. Samples were then capped and heated at 121 °C for 2 h. Afterward, samples were dried down under nitrogen at 74 °C.

Reduction of Sugar Aldehydes: A total of 475 µL of a freshly prepared 10 mg/mL solution of sodium borohydride in 1 M ammonium hydroxide was added to each test tube. After letting the sample react for 1 h at room temperature, 63 µL of methanol was added to each sample, then dried down at 74 °C under nitrogen. A solution of 9:1 (v/v) methanol: acetic acid was then prepared and 125 µL was added to each test tube, which

was again dried under nitrogen. Before moving forward, the samples were fully dried in vacuum desiccator for at least 15-20 min.

Acetylation of Nascent Hydroxyl Groups: A total of 18 µL of water was added to each sample and mixed well to dissolve the entire sample residue. A total of 250 µL of acetic anhydride was then added to each sample. Next, the sample was sonicated in a water bath for 2 min followed by an incubation for 10 min at 60 °C. 230 µL of concentrated TFA was then added to each test tube. The capped test tube was then incubated at 60°C for 10 min.

Final Cleanup: Approximately 2 mL methylene chloride was added to each test tube and mixed well. Then, 2 mL water was added to each sample and mixed well. Liquid/liquid extraction was performed twice, saving the organic layer. Next, the organic layer was transferred with a silanized glass pipette into a silanized autosampler vial. The organic layer was then evaporated under nitrogen, reconstituted in 120 µL of acetone and capped for injection onto GC-MS. A molecular overview of the global glycan methylation analysis procedure is shown in **Figure 2.2**.

Gas Chromatography-Mass Spectrometry: For sample analysis, an Agilent Model A7890 gas chromatograph (equipped with a CTC PAL autosampler) was used coupled to a Waters GCT (time-of-flight) mass spectrometer. A total of 1 µL of the sample was injected in split mode onto an Agilent split-mode liner that contained a small plug of silanized glass wool with the temperature set to 280 °C. For all samples, one injection was made at split ratio of 20:1. A 30-m DB-5 ms GC column was used for chromatography. The oven temperature was initially held at 165 °C for 0.5 min. Then, the temperature increased 10 °C/min up to 265 °C, followed by immediate increase of 30 °C/min to 325 °C where it was kept constant for 3min. Total run time was 15.5 min. The temperature of the transfer line was kept at 250 °C. After the sample components were

eluted from the GC column, they were subjected to electron ionization with an electron energy of 70 eV at a temperature of 250 °C. The *m/z* range of analysis was 40-800 with a scan cycle time of 0.1 s. Perfluorotributylamine was used for daily tuning and calibration of the mass spectrometer.

2.2.4 Data Processing

Quantification was done by integrating the summed extracted ion chromatogram peak areas (details provided elsewhere [88]), using QuanLynx software. The peaks were integrated automatically and verified manually. Then, all the information given by integration was exported to a spreadsheet for further analysis.

2.2.5 Statistical Analysis

The peak areas for individual hexoses were normalized to sum of all the endogenous hexoses and those of individual HexNAcs were normalized to sum of all endogenous HexNAcs. For the stability test, beside this normalization method, the peak areas for individual hexoses were normalized to heavy glucose and those of individual HexNAcs were normalized to heavy N-acetyl glucosamine (heavy GlcNAc). Differences between different sample matrices as well as different sample handling conditions were evaluated by means of the Friedman test followed by Dunn's post hoc test using GraphPad Prism 7.

2.3 Results and Discussion

2.3.1 Glycan Node measurements in Different Sample Matrices

Only a few statistically significant differences between the P/S matrices were observed (**Tables 2.1 and 2.2**). Based on these results, sodium citrate and sodium EDTA plasma samples, which account for all of the pair-wise differences observed in **Table 2.1**, were

excluded from clinical studies involving “glycan node analysis”. A few remaining differences (noted within the smaller sample set involving heparin plasma; **Table 2.2**), while statistically significant, were small and actually within the interassay precision range for the relevant markers [106].

2.3.2 Stability of Glycan Nodes in Different Conditions

No statistically significant difference was observed between the samples handled in conditions other than -80 °C and controls. Comparing glycans in only the EDTA plasma samples (excluding serum and Heparin plasma) at different conditions to the controls resulted in one slightly significant difference in 6-linked galactose between samples stored at 25 °C for 1 day and their controls (**Table 2.3**), which is within the interassay precision range for this marker [106].

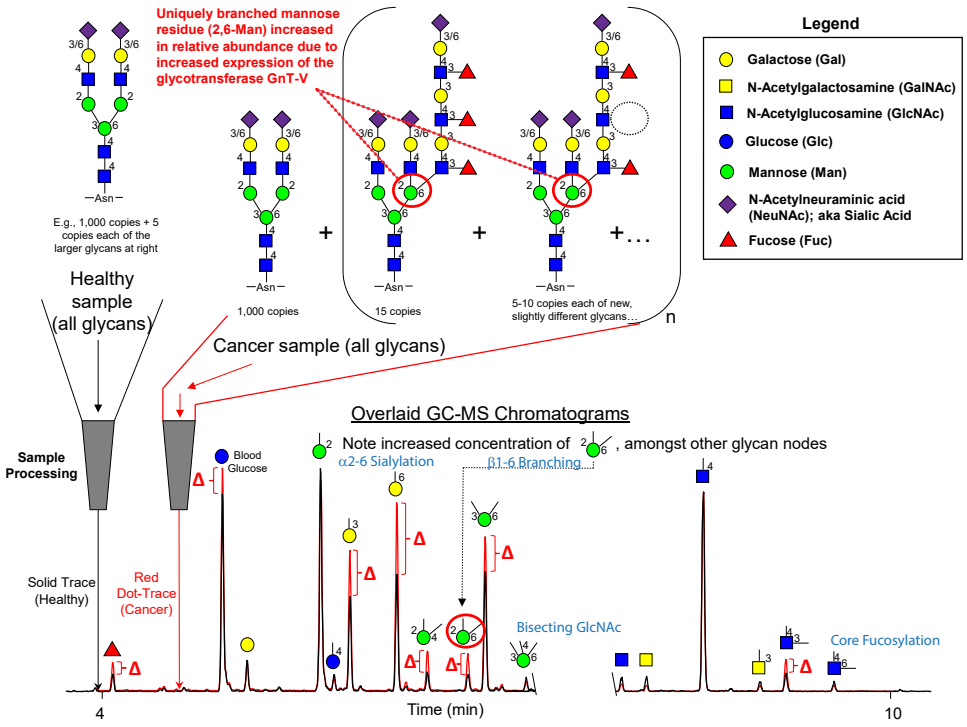

Figure 2.1. Conceptual overview of the glycan "node" analysis concept—which essentially consists of applying glycan linkage (methylation) analysis to whole biofluids. Intact normal and abnormal glycans including O-glycans, N-glycans and glycolipids, are processed and transformed into partially methylated alditol acetates (PMAAs, Fig. 2.2), each of which corresponds to a particular monosaccharide-and-linkage-specific glycan "node" in the original polymer. As illustrated, analytically pooling together the glycan nodes from amongst all the aberrant intact glycan structures provides a more direct surrogate measurement of abnormal glycosyltransferase activity than any individual intact glycan, while simultaneously converting unique glycan features such as "core fucosylation", "α2-6 sialylation", "bisecting GlcNAc", and "β1-6 branching" into single analytical signals. Actual extracted ion chromatograms from 9-µL blood plasma samples are shown. Numbers adjacent to monosaccharide residues in glycan structures indicate the position at which the higher residue is linked to the lower residue. Figure adapted with permission from Borges CR et al. Anal. Chem. 2013, 85(5):2927-2936. Copyright 2013 American Chemical Society.

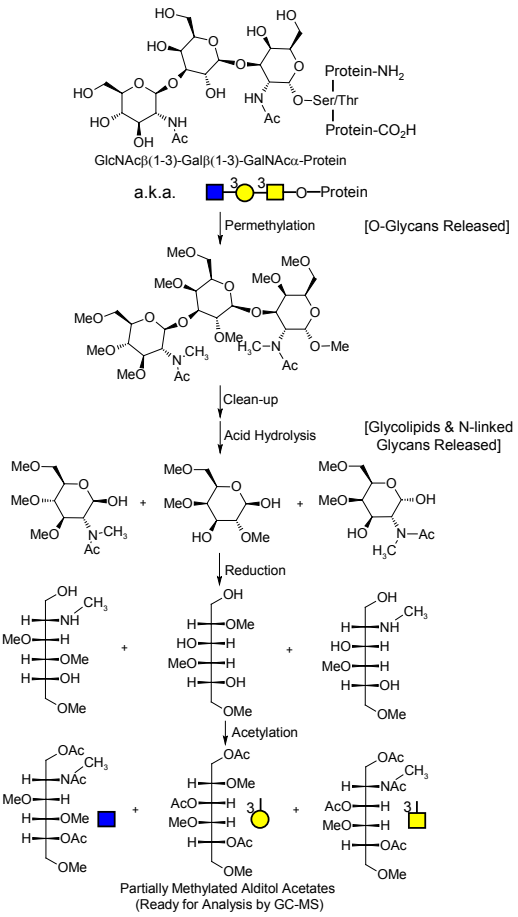

Figure 2.2. Molecular overview of the glycan "node" analysis procedure. For glycans from blood plasma and other biofluids, O-linked glycans are released during permethylation, while N-linked glycans and glycolipids are released during acid hydrolysis. The unique pattern of methylation and acetylation in the final partially methylated alditol acetates (PMAAs) corresponds to the unique "glycan node" in the original glycan polymer and provides the molecular basis for separation and quantification by GC-MS. Figure adapted with permission from Borges CR et al. Anal. Chem. 2013, 85(5):2927-2936. Copyright 2013 American Chemical Society.

Table 2.1. Impact of plasma and serum matrices on glycan node measurements. Data were acquired from 21 healthy individuals from which 5 different plasma or serum matrices were collected at the same draw. Differences between matrices were evaluated by the Friedman test followed by Dunn's post hoc test; $p > 0.05$ (ns), $p < 0.05$ (*), $p < 0.01$ (**).

Glycan Node	K_2EDTA vs. K_3EDTA	K_2EDTA vs. Na EDTA	K_2EDTA vs. 3.8% Na Citrate	K_2EDTA vs. Serum	K_3EDTA vs. Na EDTA	K_3EDTA vs. 3.8% Na Citrate	K_3EDTA vs. Serum	Na EDTA vs. 3.8% Na Citrate	Na EDTA vs. Serum	3.8% Na Citrate vs. Serum
t-Fucose	ns	ns	ns	ns	ns	ns	ns	ns	ns	*
t-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
2-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4-Glc	ns	ns	ns	ns	ns	*	ns	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,4-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
2,4-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
2,6-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,6-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,6-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,4,6-Man	ns	ns	ns	ns	ns	*	ns	ns	ns	ns
t-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	ns	ns	ns	ns	ns	ns	**	**
3,4-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4,6-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,6-GalNAc	ns	ns	**	ns	ns	ns	ns	ns	ns	**

Table 2.2. Comparison of sodium heparin plasma to potassium EDTA plasma and serum matrices. Data were acquired from 6 replicate analyses (per matrix) of a matched set of samples taken from a single donor during one blood draw. Differences between matrices were evaluated by the Friedman test followed by Dunn's post hoc test; $p > 0.05$ (ns), $p < 0.05$ (*).

Glycan Node	Heparin vs. K_2EDTA	Heparin vs. Serum	K_2EDTA vs. Serum
t-Fucose	ns	ns	ns
t-Gal	ns	ns	ns
2-Man	ns	ns	ns
4-Glc	ns	ns	ns
3-Gal	ns	ns	ns
6-Gal	*	ns	ns
3,4-Gal	*	*	ns
2,4-Man	ns	ns	*
2,6-Man	ns	*	ns
3,6-Man	ns	ns	ns
3,6-Gal	ns	ns	ns
3,4,6-Man	ns	ns	ns
t-GlcNAc	ns	ns	ns
4-GlcNAc	ns	ns	ns
3-GlcNAc	ns	ns	ns
3-GalNAc	ns	ns	ns
3,4-GlcNAc	ns	*	ns
4,6-GlcNAc	ns	ns	ns
3,6-GalNAc	ns	ns	ns

Table 2.3. Comparison of glycan node stability at different conditions to controls stored at -80 °C. Hexose data were normalized to heavy, stable isotope labeled glucose (Glc) and HexNAc data normalized to heavy, stable isotope labeled GlcNAc. Differences between time points were evaluated by Friedman test followed by Dunn's post hoc test; $p > 0.05$ (ns), $p < 0.05$ (*).

Glycan Node	10 days at -20 C	90 days at -20 C	360 days at -20 C	2 days at 4 C	90 days at 4 C	1 day at 25 C
t-Fucose	ns	ns	ns	ns	ns	ns
t-Gal	ns	ns	ns	ns	ns	ns
2-Man	ns	ns	ns	ns	ns	ns
4-Glc	ns	ns	ns	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	ns	ns	ns	*
2,4-Man	ns	ns	ns	ns	ns	ns
2,6-Man	ns	ns	ns	ns	ns	ns
3,6-Man	ns	ns	ns	ns	ns	ns
3,4,6-Man	ns	ns	ns	ns	ns	ns
t-GlcNAc	ns	ns	ns	ns	ns	ns
4-GlcNAc	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	ns	ns	ns	ns
3,4-GlcNAc	ns	ns	ns	ns	ns	ns
4,6-GlcNAc	ns	ns	ns	ns	ns	ns

CHAPTER 3

STAGE DEPENDENCE, CELL-ORIGIN INDEPENDENCE, AND PROGNOSTIC CAPACITY OF SERUM GLYCAN FUCOSYLATION, β1-4 BRANCHING, β1-6 BRANCHING AND α2-6 SIALYLATION IN CANCER [2]

3.1 Introduction

To date, we have only reported results from pilot studies in which our molecularly bottom-up glycomics methodology known as "glycan nodes analysis" was applied to (mostly) advanced stages of lung [88] and breast cancer [107]. In order to gain a representative perspective on the potential utility of this approach to detecting a variety of different types of cancer at varying stages, we have now applied it to over 950 clinical P/S samples from 7 different case control studies across all stages of cancer in which the cancer cases were compared to related benign conditions and/or healthy controls. A study of plasma samples from 428 Stage I-IV lung cancer patients, age/gender/smoking-status matched controls, and certifiably healthy living kidney donors served as the backbone for this report—in which plasma from a single donor served as a quality control specimen in every single batch of samples—facilitating comparisons to pancreatic (rapid autopsy), ovarian (Stage III), prostate (Stage II), and a large independent lung cancer (Stage I) case-control study. Based on the behavior of P/S glycans established to date, we hypothesized that the alteration of P/S glycans observed in cancer would be independent of the tissue in which the tumor originated yet exhibit stage dependence that varied little across cancers classified on the basis of tumor origin.

[2] Reprinted with permission from S. Ferdosi *et al.*, "Stage Dependence, Cell-Origin Independence, and Prognostic Capacity of Serum Glycan Fucosylation, β1–4 Branching, β1–6 Branching, and α2–6 Sialylation in Cancer," *J. Proteome Res.*, p. acs.jproteome.7b00672, 2018. Copyright 2018, American Chemical Society.

3.2 Materials and Methods

3.2.1 Materials

Materials for glycan node analysis are listed in section 2.2.1. Glucose colorimetric assay kit was obtained from Cayman Chemical (Cat. No. 10009582).

3.2.2 Plasma and Serum Samples

A summary of the case-control sample sets employed in this study is provided in **Table 3.1**. All specimens were collected in compliance with the Declaration of Helsinki principles. Once collected they were coded and de-identified to protect patient identities.

Living Kidney Donors: EDTA plasma samples from certifiably healthy living kidney donors were enrolled in the Multidisciplinary Biobank at Mayo Clinic Arizona under a Mayo Clinic IRB approved protocol. Patients eligible for enrollment were those seen at Mayo Clinic Arizona who were ≥ 18 years old, able to provide informed consent, and undergoing evaluation as a potential living kidney donor. Detailed inclusion & exclusion criteria for these patients are provided in Appendix A. None of these patients smoked at the time of health screening and blood collection, 27% were former smokers and 73% never smoked. Specimens were collected over a 2-year period from Dec. 2013 to Dec. 2015. Standard operating protocols and blood collections were performed as previously described [108]. All specimens were stored at -80 °C prior to shipment to ASU.

Large Lung Cancer Set: Sodium heparin plasma samples for the large lung cancer study were collected at the University of Texas MD Anderson Cancer Center under the supervision of Dr. Xifeng Wu. Heparin is a glycosaminoglycan itself but the vast majority of its monomer units are carboxylated and/or sulfated. As we have previously described [88], sulfated and carboxylated glycan monomers cannot be detected by the analytical

methodology employed in this study. The PMAA from 4-linked GlcNAc could theoretically be produced by the heparin anticoagulant, but empirically we found in our matched collection studies (described in Chapter 2) that 4-linked GlcNAc from heparin plasma is not significantly different from EDTA plasma or serum. Specimens for lung cancer cases and controls from the University of Texas MD Anderson Cancer Center included in this study are part of an ongoing large lung cancer study that has been recruiting since 1995. This study has received approval from the University of Texas MD Anderson Cancer Center and Kelsey-Seybold institutional review boards. Venous blood was drawn from newly-diagnosed and histologically confirmed lung cancer patients (prior to therapy) and age/gender/ethnicity-matched controls at the MD Anderson Cancer Center hospital and the nearby Kelsey Seybold Clinic, respectively. All blood was drawn and processed under the same SOP. Patients were not necessarily in a fasted state. Blood was centrifuged then aliquoted and placed into a liquid nitrogen tank. After collection samples were coded and de-identified prior to shipment to Arizona State University for analysis. A more detailed profile of the clinical characteristics of the patients in this large lung cancer study is provided in **Table 3.2**.

Liver Fibrosis (Non-Cancerous): Serum samples from patients at all stages of liver fibrosis were collected at the Sunnybrook Health Sciences Centre, under the direction of Dr. Lei Fu and Dr. David E. C. Cole. This study was approved by Research Ethics Board, Sunnybrook Health Sciences Centre, Toronto. Patients were recruited between 2007 and 2011. Written informed consent was obtained from each participant. All subjects with various chronic liver diseases were considered eligible if they would have liver biopsy for the diagnosis of liver fibrosis as part of their routine care. Blood specimens were collected and serum was separated from cells following standard clinical

laboratory procedures. Serum aliquots were stored in -70 °C. The specimens were coded and de-identified according to the study protocol.

Stage I Lung Adenocarcinoma: Serum samples from stage I lung adenocarcinoma patients and age/gender/smoking-status matched controls were collected under NYU Institutional Review Board (IRB) approval at the NYU Langone Medical Center by Dr. Harvey Pass. Arterial blood samples were collected from fasting patients undergoing surgery in the time frame from September 2006 to August 2013 to remove one or more lung nodules that were detected during a CT scan. Determination of whether nodules were benign or malignant was made following a pathological exam of the excised nodules. Serum was collected in standard glass serum tubes and allowed to sit upright for 30-60 minutes to allow clotting. Subsequently, tubes were centrifuged at 1,200 x g for 20 min at room temperature, then aliquoted and placed at -80 °C within 2-3 hours of collection. No freeze-thaw cycles occurred prior to shipment to Arizona State University (Borges lab) for analysis.

The original set of serum samples for analyzing stage I lung adenocarcinoma were including cases that were arterial blood samples collected from fasting patients during surgery by Dr. Harvey Pass. The age, gender, and smoking-matched controls were venous blood samples collected from patients during doctor visit by Dr. Bill Rom. This set of samples was excluded from this study because of the reasons explained in the Discussion section.

Stage II Prostate Cancer: Serum samples from stage II prostate cancer patients were obtained from the Cooperative Human Tissue Network (CHTN)—an NIH-sponsored biospecimen collection agency. The quality management system of the CHTN is described elsewhere [109]. Age-matched control samples from nominally healthy male donors were obtained from ProMedDx (Norton, MA).

Stage III Serous Ovarian Cancer: Serum specimens from stage III serous ovarian cancer patients were collected at Brigham and Women's Hospital under IRB approval by Dr. Daniel Cramer. Sera were obtained at the time of presentation prior to surgery. Age, gender, and location matched control sera from women without a history of cancer (other than non-melanoma skin cancer) were obtained from the general population under a standardized serum collection protocol. All serum samples were collected from 2001 to 2010 and were stored at -80 °C prior to analysis. These specimens have previously been described [110], [111].

Stage IV Lung Cancer: A set of serum samples from stage IV lung cancer patients and age/gender matched nominally healthy control donors that was completely separate from those provided by Dr. Xifeng Wu at the University of Texas MD Anderson Cancer Center was obtained from ProMedDx.

Rapid Autopsy Pancreatic Cancer: Serum specimens from rapid autopsy patients who had recently died from pancreatic cancer were collected by Dr. Michael Hollingsworth at the University of Nebraska Medical Center under IRB approval. These samples have previously been described [112]. In brief, specimens were collected within 2-3 hrs of death. Control serum samples were from patients with benign pancreatic conditions and elevated CA19-9 levels. Samples were coded, de-identified and kept at -80 °C prior to shipment to ASU.

Additional Biospecimen Details: As described above, all blood samples were processed into P/S immediately following collection and stored at -70 °C or colder until analyzed. Following shipment in dry ice, vial headspace was vented prior to thawing to avoid CO_2-mediated sample acidification [113]. The molecular integrity of the sample set that showed the greatest differences between cases and controls (rapid autopsy pancreatic cancer sera) was examined using an assay based on *ex vivo* protein oxidation

that was recently developed by the Borges group [114]. The prostate cancer and stage I lung adenocarcinoma sets were spot-checked as well. No samples produced evidence for concern about specimen integrity.

In this study, multiple independent sets of sample were compared to each other. Each case-control set was analyzed blind and in random order. Within each batch, across all sets, a quality control (QC) EDTA plasma sample was included consisting of a 9-uL aliquot of the same bulk plasma sample in order to verify the reproducibility across batches. Notably, the samples from the certifiably healthy living kidney donors were analyzed in separate batches of samples from those in the large lung cancer set. To justify direct comparison of these two sets of samples, we verified that the average values measured for each glycan node in the two sets of QC sample results were not statistically significantly different. Moreover, if the average value of the QC sample was slightly higher or lower in the large lung cancer set relative to the living kidney donor set a scaling factor based on this difference in QC samples was employed to adjust the living kidney donor data set. For each glycan node, this adjustment brought the living kidney donor data set distribution slightly closer to the control distributions observed in the large lung cancer set—meaning that it was a conservative adjustment.

3.2.3 Experimental Procedures

3.2.3.1 Glycan Node Analysis

The experimental procedure and data processing for glycan node analysis are described in details in sections 2.2.3 and 2.2.4, respectively.

3.2.3.2 Glucose Oxidase-Based Assay

Glucose colorimetric assay kit was used for measuring the concentration of glucose in serum samples. A total of 10 mL of 250 mM sodium phosphate assay buffer was diluted with 40 mL of HPLC-grade water in order to be used for dilution of glucose standards and enzyme mixture. To prepare standard solutions a 100 mg/dL stock solution was made using 50 µL of the 1000 mg/dL glucose standard and 450 µL of diluted assay buffer. Using the stock solution, glucose standards with concentrations of 0, 2.5, 5, 7.5, 10, 15, 20, and 25 mg/dL were made. To perform the assay, 15 µL of each standard and sample was added to two wells followed by 85 µL of diluted assay buffer. By adding 100 µL of enzyme mixture, which was reconstituted with 6 mL of diluted assay buffer, the reaction was initiated. Then, the plate was covered with the plate cover and incubated for 10 min at 37 °C. The plate cover was then removed and the absorbance was read at 500-520 nm using a Thermo Scientific Multiskan Go plate reader. The concentrations of glucose in the samples were calculated using SkanIt Software 3.2.

3.2.4 Statistical Analysis

The peak area for each glycan node was normalized in one of two possible ways: In the first approach, individual hexoses were normalized to heavy glucose and individual HexNAcs were normalized to heavy N-acetyl glucosamine (heavy GlcNAc). (Notably, these two internal standards were omitted during analysis of the prostate cancer set of samples.) In the second approach, individual hexoses were normalized to the sum of all endogenous hexoses and individual HexNAcs were normalized to the sum of all endogenous HexNAcs. This normalization scheme provided modestly improved within-batch reproducibility, but limited observation of potential simultaneous increases in all glycan nodes. Each stage of each cohort was log-transformed and outliers were removed

with the ROUT method at Q = 1% using GraphPad Prism 7. Data were then reversed transformed by taking the anti-log of each value. Differences between patient cohorts and stages in the large lung cancer study were evaluated by means of the Kruskal-Wallis test followed by the Benjamini-Hochberg false discovery correction procedure using R version 3.3.3. This software was also used to generate and plot receiver operating characteristic (ROC) curves. ROC curves were compared via DeLong's test using RStudio Version 1.0.143. Stage-by-stage multivariate modeling on the large lung cancer set was carried out using multivariate logistic regression, with performance assessed by leave-one-out cross-validation, and model selection carried out using a best subsets procedure. These analyses were carried out using R version 3.3.3. The ability of particular glycan nodes to predict cancer progression and survival was assessed via Cox proportional hazards regression models using XLSTAT Version 2012.3.01—the results of which were verified (duplicated) using SAS 9.4. Survival curves were generated and associated log-rank Mantel-Cox tests were carried out using GraphPad Prism 7.

3.3 Results

The primary focus of this study was the large lung cancer set as it constituted the single largest set and covered all stages of cancer. A summary of all sample sets analyzed as part of this study is provided in **Table 3.1**. In total, 19 glycan "nodes" were measured with relative abundances that were consistently greater than 1% of respective total hexoses or total N-acetylhexosamines (HexNAcs). As reported elsewhere, this threshold ensures quantitative precision between batches of samples [88], [106]. Relative to the age/gender/smoking-status matched controls, significant changes were observed in 4 out of 19, 2 out of 19, 17 out of 19 and 17 out of 19 nodes in plasma samples taken from stage I, II, III and IV patients, respectively (**Table 3.3**). Based on normalization to

heavy, stable isotope-labeled glucose and GlcNAc internal standards, all altered glycan nodes except 4-Glc (which is mostly derived from glycolipids) were elevated in the cancer patients relative to the controls. Analogous results for data in which each hexose was normalized to the sum of endogenous hexoses and each HexNAc was normalized to the sum of endogenous HexNAcs revealed that this alternate normalization procedure is not as effective at teasing out differences between the cohorts in the large lung cancer study (**Table 3.4**).

3.3.1 Highly Altered Glycan Features

The five glycan nodes that were most elevated in the cancer cases relative to the at-risk controls included the following: 1) Terminal fucose—which corresponds to essentially all fucose in blood plasma. (Non-terminal fucose is only found in Notch proteins [88], [3] which, at most, would contribute only an infinitesimal fraction of the fucose found in blood plasma and, if ever detected by the approach employed here, would be observed as 3-linked fucose.) 2) 6-linked galactose, which corresponds specifically to α2-6 sialylation and almost completely to the activity of the ST6GalI glycosyltransferase enzyme [88]; 3) 2,4-linked mannose, which corresponds to β1-4 branching of N-linked glycans and almost completely to the activity of the GnT-IVa enzyme [88]; 4) 2,6-linked mannose, which corresponds to β1-6 branching of N-linked glycans and to the activity of the GnT-V enzyme [88]; 5) 3,4-linked N-acetylglucosamine (GlcNAc), which predominately corresponds to outer-arm fucosylation and the activity of the FucT-III, FucT-V, FucT-VI, and FucT-XI enzymes [88]. The univariate distributions of these five glycan nodes (normalized to heavy glucose or heavy GlcNAc added as an internal standard), along with receiver operating characteristic (ROC) curves that describe the potential clinical relevance of their distributions are shown in **Figure 3.1**.

3.3.2 Stage and Health-Status Dependence

Table 3.3 and **Figure 3.1** illustrate both the strong stage-dependence of these glycan features as well as the notable contrast of their distributions in certifiably healthy individuals compared to the general middle-aged to elderly population (i.e., "controls") who are at a similar risk for cancer as individuals who actually had cancer. Similar distributions and trends were noted when the five glycan nodes were normalized to the sum of endogenous hexoses or HexNAcs, but the ROC c-statistics (areas under the curve, AUCs) tended not to be as large (**Figure 3.2**). The average age of the certifiably healthy living kidney donor population was 47 and that of both the controls and lung cancer cases in this set of samples was 61 (**Table 3.2**). However, after pooling data from the certifiably healthy donors, controls and lung cancer cases and correcting for multiple comparisons, no statistically significant correlations with age were observed for any of these glycan nodes. (Before correcting for multiple comparisons, terminal (total) fucose appeared slightly correlated with age (Pearson correlation $R^2 = 0.013$ and $p = 0.021$), but this result cannot be considered statistically significant after considering the fact that multiple comparisons were made.) Likewise, no statistically significant correlations of glycan nodes with age were observed when these groups of patients were evaluated individually.

In general, the five glycan nodes increased together as the stage of cancer advanced (**Figure 3.3**). Moreover, the behavior of these nodes was independent of the organ of tumor origin, at least when comparing lung cancer with pancreatic, ovarian and prostate cancers (**Table 3.5**). This is evident when panels a, c and d of the left column of **Figure 3.3** are compared with their respective lung cancer stages shown in the right column. At stage IV both outer-arm fucosylation and terminal (total) fucosylation lag a bit behind α2-6 sialylation, β1-4 branching and β1-6 branching—but fucosylation-related

nodes caught up and even surpassed these other glycan features once cancer had fully run its course (**Figure 3.3a**). When stage was held constant, no glycan nodes were found to be significantly different between adenocarcinoma, squamous cell carcinoma and small cell carcinoma—the three different histological sub-types observed in the large lung cancer study (**Table 3.2**). However, terminal (total) fucosylation, β1-6 branching, and outer-arm fucosylation were altered within the control cohort on the basis of smoking status (grouped as never-smokers, former smokers or current smokers; **Figure 3.4**).

Cases and controls in the original set of stage I lung adenocarcinoma patients were best separated when data were normalized to terminal glucose (t-Glc) (**Figure 3.5**). In order to determine if the source of these differences was blood glucose or the glycan nodes, a glucose oxidase-based assay was used to measure the amount of free glucose in these serum samples. This glucose assay revealed that the cases had significantly higher level of glucose concentration with the average of 92.23 mg/dL relative to controls with that of 77.52 mg/dL (Mann Whitney test; $p < 0.0001$).

3.3.3 Orthogonality of Glycan Features

In order to evaluate the orthogonality of all 19 glycan nodes included in this study (**Table 3.3**), multivariate logistic regression models were created for the large lung cancer set on a stage-by-stage basis (**Figure 3.6**). Results of modeling are shown as ROC curves, where the model-derived predicted probability of disease for the sample was used as the discriminatory variable. In summary, the fully cross-validated forms of these multivariate logistic regression models do not distinguish lung cancer cases from controls any better than individual glycan nodes (cf. **Figure 3.3f-i**). This indicates a

general lack of orthogonality or independence between the glycan features observed in this study.

3.3.4 Comparison to Liver Fibrosis

The vast majority of glycoproteins found in blood P/S are derived from either liver glycoproteins or immunoglobulins (IgG molecules) secreted by the immune system [115], [116]. Essentially all non-protein targeting serum glycomics approaches, including the one employed in this study, detect changes in these glycans and not novel glycans secreted or sloughed-off by cancer cells. This concept has been acknowledged elsewhere [117]. Nevertheless, P/S glycans are notoriously known for being altered in cancer [77], [90]–[92], [118]. However, they are also known to be altered in inflammatory conditions in the absence of cancer [119]–[121]. As an initial attempt to begin to parse out the behavior of the five glycan nodes that were most elevated in the large lung cancer set, they were analyzed in a set of serum samples from liver fibrosis patients (**Figure 3.7**). Statistical analysis (Kruskal-Wallis) indicated that there were no significant differences in any of the glycan nodes shown across all stages of liver fibrosis. This may have been due to limited statistical power. Notably, however, fucosylation-related markers exhibited a tendency to be elevated in stage III-IV liver fibrosis.

3.3.5 Prediction of Progression and All-Cause Mortality

The five glycan nodes that were most elevated in the large lung cancer set were evaluated for their ability to predict both progression and all-cause mortality in a Cox proportional hazards regression model. After adjusting for age, gender, smoking status and cancer stage, only 6-linked galactose, which corresponds to α2-6 sialylation, predicted both progression and all-cause mortality with p-values of < 0.01 when the glycan nodes were

modeled as continuous variables. All four other top-performing glycan nodes were able to predict survival ($p < 0.05$), but only β1-4 branching and β1-6 branching were also able to predict progression ($p < 0.05$). Because relative rather than absolute quantification was employed, glycan node units lack readily interpretable meaning. As such, measurements of α2-6 sialylation were broken into quartiles and the Cox proportional hazards analysis repeated. After adjusting for age, gender, smoking status and cancer stage, the top α2-6 sialylation quartile predicted progression with a hazard ratio of 2.45 relative to all other quartiles combined (lower bound at 95% CL = 1.54; upper bound at 95% CL = 3.90; $p = 1.5 \times 10^{-4}$). Likewise, after the same adjustments, the top α2-6 sialylation quartile predicted all-cause mortality with a hazard ratio of 1.52 relative to all other quartiles combined (lower bound at 95% CL = 1.02; upper bound at 95% CL = 2.23; $p = 0.042$). Progression and survival curves illustrate the differences in the rates of occurrence of these events for the top α2-6 sialylation quartile vs. all other quartiles (**Figure 3.8**). Progression and survival curves for stage III patients alone illustrate that the separation of progression by α2-6 sialylation and the separation of survival by α2-6 sialylation is not simply driven by stage (**Figure 3.8c-d**). The α2-6 sialylation was not elevated or able to predict progression or survival in the stage I lung adenocarcinoma set.

3.4 Discussion

The five glycan features that were most elevated relative to healthy individuals and at-risk controls were terminal (total) fucosylation, α2-6 sialylation, β1-4 branching, β1-6 branching and outer-arm fucosylation (**Table 3.3** and **Figures 3.1 and 3.3**). Two phenomena stood out most with regard to their distributions amongst the cohorts of the large lung cancer study: First, there was a striking stage dependence of all five glycan features that was independent of the tumor organ of origin (**Table 3.3**, **Figures 3.1**

and 3.3, and **Table 3.5**). In part, statistical significance at earlier stages may not have been achieved due to the relatively low number of samples measured from patients at stages I-II (n ~ 20 per stage). Statistically significant elevation of core-branched O-glycans (i.e., 3,6-linked GalNAc) over age/gender/smoking-status matched controls was observed in stage I lung adenocarcinoma from this separate, larger set of samples (**Figure 3.3**). However, it was clear from the ROC curve (**Figure 3.3e**) that this glycan node cannot serve as a useful early stage diagnostic biomarker. The original set of stage I lung adenocarcinoma samples was excluded from this study for two reasons: 1) Since the samples from cancer patients were collected during surgery, these patients had a significantly higher level of glucose (which was caused by pre-surgical hyperglycemia) relative to controls that were collected from patients during doctor visit. This difference between the location/setting of blood collection explained the elevated glucose in the case samples measured by the glucose oxidase-based assay. 2) After analyzing this set of samples, it was disclosed by our collaborators that the controls were exposed to the thawed state significantly longer than cases. Although based on the stability test this second issue does not have a major impact on glycan node stability, it can be a major issue for other analytes. In general, the importance of these two problems could easily be reversed, depending on the analyte(s) of interest. The secondary set of stage I lung adenocarcinoma, when normalized to the sum of endogenous hexoses or hexNAcs, demonstrated similar ROC curves to the original set for the top performing glycan nodes (**Figure 3.5a,c**). However, when normalized by t-Glc, they are significantly different (**Figure 3.5b,d**). The separation of cases and controls in the original set of stage I lung adenocarcinoma is caused by the elevated levels of glucose in cancer patients, which is not clinically valuable.

A second notable feature apparent in the large lung cancer set was the statistically significant difference between certifiably healthy living kidney donors and risk-matched controls for α2-6 sialylation, β1-4 branching and β1-6 branching—with controls always increased toward the direction of cancer (**Table 3.3** and **Figure 3.1**). These differences between certifiably healthy individuals and patients with an elevated risk of cancer underscore the high risk of false discovery when nominally healthy sample donors rather than well-characterized, clinically relevant controls are employed during biomarker development. The notable differences between healthy individuals and at-risk controls also supports the idea that the biological landscape within plasma/serum may undergo "grooming", "conditioning" or pre-metastatic "niche" formation prior to cancer taking hold within the body [122]–[126]. Given that inflammation is closely tied to the development of cancer [127], [128] and that at least some glycans and glycan features are known to be altered in inflammatory conditions in the absence of cancer [119]–[121], pre-cancerous inflammation may be responsible for the elevation of many of the glycan features observed in the at-risk controls relative to the certifiably healthy living kidney donors—suggesting that the goal of preventing such a pre-cancerous state may be as important as preventing the transition from an at-risk state to stage I cancer. With this in mind, it is interesting to note that about 62% of the age-qualified U.S. population would be excluded as living kidney donors due to preventable health conditions.

A few studies have been published that are closely related to the one reported here, but in which intact glycans were analyzed [80], [117], [129]. While not in conflict with any of these studies, our most prominent findings of increased terminal (total) fucosylation, α2-6 sialylation, β1-4 branching, β1-6 branching and outer-arm fucosylation in stage III-IV lung cancer are most closely aligned with the major changes reported by Vasseur et al [117] for intact glycans in lung cancer. They reported

significant increases in fucosylated tri- and tetra-antennary structures, outer-arm fucosylated structures, and α2-6 sialylated structures. Moreover, they reported that all of these features were elevated in control-group former smokers relative to control group non-smokers. We found increases in terminal (total) fucosylation, β1-6 branching and outer-arm fucosylation in current smokers relative to never smokers, but only increases in terminal (total) fucosylation and outer-arm fucosylation in former smokers relative to never smokers (**Figure 3.4**). Notably, the methods employed for the analysis of intact glycans by Vasseur et al is one of just a few approaches that are capable of distinguishing 6-linked from 3-linked sialic acid [130]–[132].

Multivariate logistic regression models were not able to outperform individual glycan nodes (cf **Figure 3.3f-i** and **Figure 3.6**) with regard to distinguishing cancer at stages I-IV from controls. This indicates a general lack of biological orthogonality amongst the abnormal glycan features observed—suggesting that they likely have a singular (or small set of closely related) upstream causes: The concentration of glycoproteins in P/S is in the tens of milligrams per milliliter. As such, the observations of significant changes in P/S glycans observed here cannot be due to glycoproteins shed directly from cancer cells; almost certainly they are derived from alterations to one or both of the two major sources of P/S glycoproteins—namely liver glycoproteins or immunoglobulins (IgG molecules) secreted by the immune system [115], [116]. Such alterations are thought to be mediated by cytokines secreted from the tumor microenvironment and exist as part of an acute-phase inflammatory response [117], [133]–[137].

However, this is not to imply that liver glycoprotein and immunoglobulin glycan alterations are unimportant or lack a cancer-relevant pathological effect. Several cancer-upregulated glycoforms that cancer cells have in common with glycans that are induced

on acute phase liver proteins and/or IgG molecules in the presence of cancer have been found to mediate specific immune-modulating effects—some of which overtly favor cancer progression:

Galectins are a family of lectins that bind β-galactoside sugars within glycans and are known to modulate a variety of immunological processes involved in cancer [125], [138], [139]. Malignant T-cells in mycosis fungoides/Sezary syndrome have been found to resist galectin-1 mediated apoptosis because they both lack the CD7 receptors that carry the oligosaccharides recognized by galectin-1 and because they express sialylated core 1 O-glycans that promote galectin-1 resistance [140]. Poly-N-acetylactosamine-modified core 2 O-glycans bind to galectin-3, reducing the affinity of tumor major histocompatibility complex (MHC) class I-related chain A (MICA) for the activating NKG2D receptor on natural killer (NK) cells, preventing tumor cell killing of core 2 O-glycan expressing cancer cells [141]–[144]. Similarly, modification of MUC1 by poly-N-acetylactosamine and subsequent binding by galectin-3 interferes with TRAIL-mediated killing of DR4-expressing cancer cells by NK cells [144]–[146]. But perhaps the best known example is the ability of excessive tumor cell surface sialylation to continually stimulate the inhibitory Siglec-7 receptor on NK cells, preventing their activation [144], [147]–[149].

In light of these discoveries, the fact that α2-6 sialylation of abundant plasma/serum proteins is both associated with metastasis and poor prognosis [150], [151] and, in our study, was not only elevated in lung cancer but predicted progression and all-cause mortality in the large lung cancer set may shed additional light on a means by which cancer potentially manipulates the immune system to groom the physiological landscape and carve out a metastatic niche: Rather than directly interacting (cell-to-cell) with NK cells, tumor cells may simply be able to send out cytokine signals that are picked

up by the liver and/or the immune system that alter the way that these nominally healthy tissues glycosylate their secreted proteins. This could, for example, facilitate a large-scale amplification of sialylated glycans that are able to continually activate Siglec-7 receptors on NK cells, preventing them from killing tumor cells and allowing them to metastasize. The possibility that cancer cells may induce the abnormal glycosylation of the highly abundant liver glycoproteins and/or IgG molecules found in P/S as a shielding mechanism against innate immune detection during metastasis attempts has received very little attention, but may be worth investigating. Though speculative, this strategy could even potentially be deployed in cases where cancer cells deplete themselves of a glycan feature required for immune-cell recognition—such as fucosylation recognized by the TRAIL-mediate killing mechanism of NK cells [152]—but induce it on abundant P/S proteins, serving to "swamp out" the recognition mechanism of innate immune surveillance.

The ability of α2-6 sialylation to predict lung cancer progression and survival is not unique among P/S glycans. Indeed, all five top-performing glycan nodes in the present study were able to predict progression and/or survival to a more limited extent than α2-6 sialylation. The prognostic capacity of β1-4 and β1-6 branching however, may, at least in part, be due to the fact that these glycan features simply create greater opportunity for sialylation. Beyond this study, others have found that the sialyl Lewis X epitope (which displays α2-3 sialylation rather than α2-6 sialylation) predicts progression and survival in both small cell [153] and non-small cell lung cancer [154]–[156]. Like the prognostic Veristrat markers [157]–[159], which are serum amyloid A proteoforms [160], elevated α2-6 sialylation in lung cancer may largely be due to an inflammatory response by the liver. But if, as described above, sialylation-based cloaking of tumor cells from the immune system plays an important role in the metastatic

process, α2-6 sialylation may turn out to play a causative, mechanistic role in lung cancer progression.

3.5 Conclusions

A molecularly bottom-up approach to plasma/serum (P/S) glycomics based on glycan linkage analysis that captures unique glycan features such as α2-6 sialylation, β1-6 branching and core fucosylation as single analytical signals was employed to evaluate the behavior of P/S glycans in all stages of lung cancer and across various stages of prostate, ovarian and pancreatic cancers. Elevation of terminal (total) fucosylation, α2-6 sialylation, β1-4 branching, β1-6 branching and outer-arm fucosylation markers were most pronounced in lung cancer in a stage-dependent manner, but these changes were found to be independent of the tumor tissue-of-origin. Using a Cox proportional hazards regression model, the marker for α2-6 sialylation was found to predict both progression and all-cause mortality in lung cancer patients after adjusting for age, gender, smoking status and stage at which the sample was taken. Interestingly, certifiably healthy P/S donors had markedly lower levels of α2-6 sialylation, β1-4 branching and β1-6 branching relative to cancer risk-matched controls. While early detection is ideal, the information provided by this and related studies [117], [119]–[121], [127], [128], [133]–[137] suggests that pre-cancerous inflammation may be responsible for the elevation of many of the glycan features observed in the at-risk controls relative to the certifiably healthy donors—implying that the goal of preventing such a pre-cancerous state may be as important as preventing the transition from an at-risk state to stage I cancer.

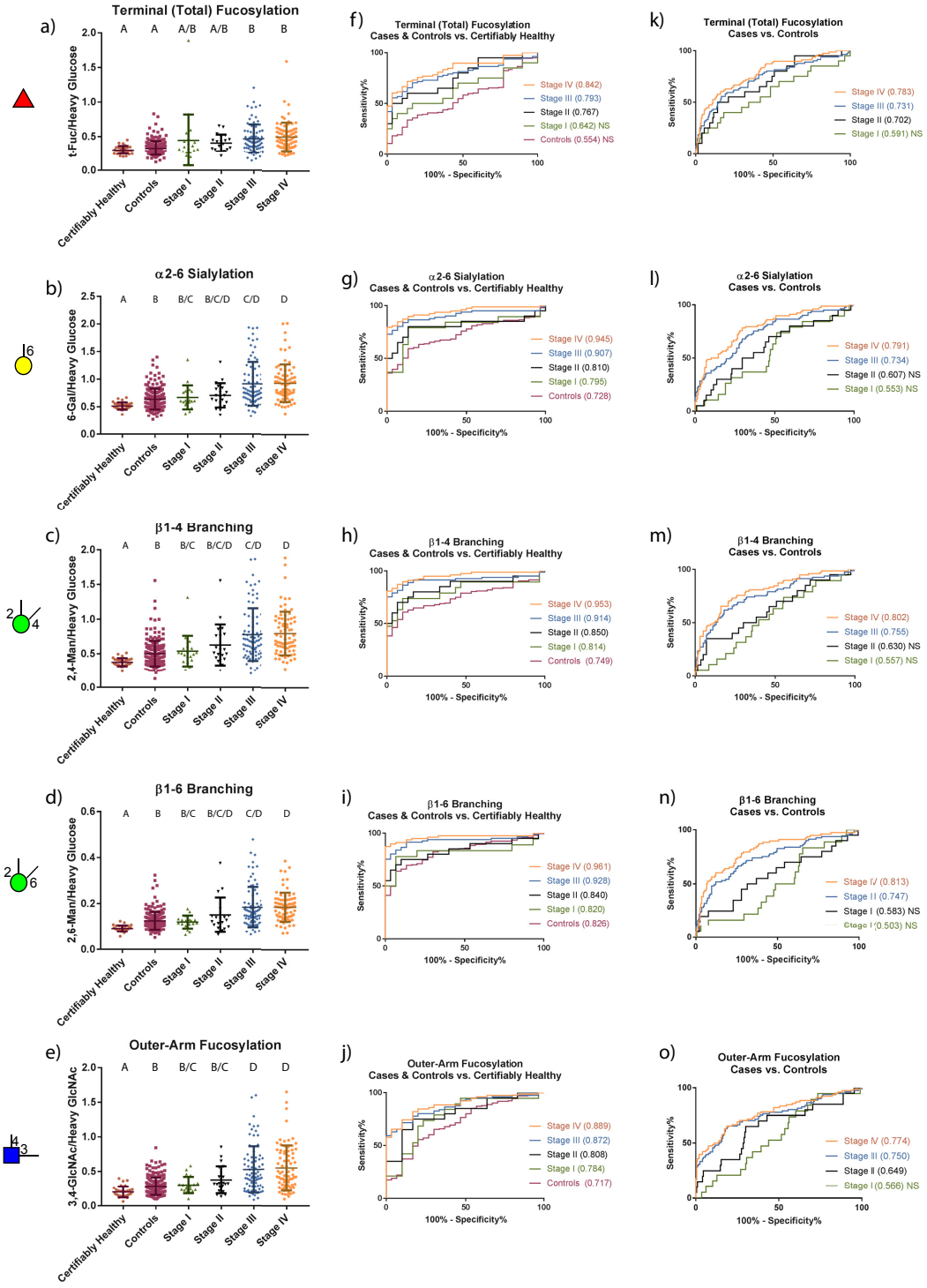
a)
Terminal (Total) Fucosylation
t-Fuc/Heavy Glucose
A A A/B A/B B B
Certifiably Healthy
Controls
Stage I
Stage II
Stage III
Stage IV
b)
α2-6 Sialylation
6-Gal/Heavy Glucose
A B B/C B/C/D C/D D
c)
β1-4 Branching
2,4-Man/Heavy Glucose
A B B/C B/C/D C/D D
d)
β1-6 Branching
2,6-Man/Heavy Glucose
A B B/C B/C/D C/D D
e)
Outer-Arm Fucosylation
3,4-GlcNAc/Heavy GlcNAc
A B B/C B/C D D
f)
Terminal (Total) Fucosylation
Cases & Controls vs. Certifiably Healthy
Stage IV (0.842)
Stage III (0.793)
Stage II (0.767)
Stage I (0.642) NS
Controls (0.554) NS
g)
α2-6 Sialylation
Cases & Controls vs. Certifiably Healthy
Stage IV (0.945)
Stage III (0.907)
Stage II (0.810)
Stage I (0.795)
Controls (0.728)
h)
β1-4 Branching
Cases & Controls vs. Certifiably Healthy
Stage IV (0.953)
Stage III (0.914)
Stage II (0.850)
Stage I (0.814)
Controls (0.749)
i)
β1-6 Branching
Cases & Controls vs. Certifiably Healthy
Stage IV (0.961)
Stage III (0.928)
Stage II (0.840)
Stage I (0.820)
Controls (0.826)
j)
Outer-Arm Fucosylation
Cases & Controls vs. Certifiably Healthy
Stage IV (0.889)
Stage III (0.872)
Stage II (0.808)
Stage I (0.784)
Controls (0.717)
k)
Terminal (Total) Fucosylation
Cases vs. Controls
Stage IV (0.783)
Stage III (0.731)
Stage II (0.702)
Stage I (0.591) NS
l)
α2-6 Sialylation
Cases vs. Controls
Stage IV (0.791)
Stage III (0.734)
Stage II (0.607) NS
Stage I (0.553) NS
m)
β1-4 Branching
Cases vs. Controls
Stage IV (0.802)
Stage III (0.755)
Stage II (0.630) NS
Stage I (0.557) NS
n)
β1-6 Branching
Cases vs. Controls
Stage IV (0.813)
Stage III (0.747)
Stage II (0.583) NS
Stage I (0.503) NS
o)
Outer-Arm Fucosylation
Cases vs. Controls
Stage IV (0.774)
Stage III (0.750)
Stage II (0.649)
Stage I (0.566) NS
Sensitivity%
100% - Specificity%

Figure 3.1. Univariate distributions and associated ROC curves for the top five-performing glycan nodes in the large lung cancer set. Letters above the data points in panels a-e indicate statistically significant differences between the six groups shown: any overlap in lettering between groups indicates a lack of significant difference between the groups (Kruskal-Wallis with Dunn's post hoc test). ROC curves for lung cancer cases (separated by stage) and controls vs. certifiably healthy patients are shown in panels f-j; stage I-IV cancer patients vs. controls are shown in panels k-o. ROC curve AUCs are provided in parenthesis next to the specified stages. "NS" next to ROC curve AUCs indicates that the ROC curve does not show a statistically significant difference between the two groups being compared. Glycan node symbol definitions are the same as in Figure 2.1.

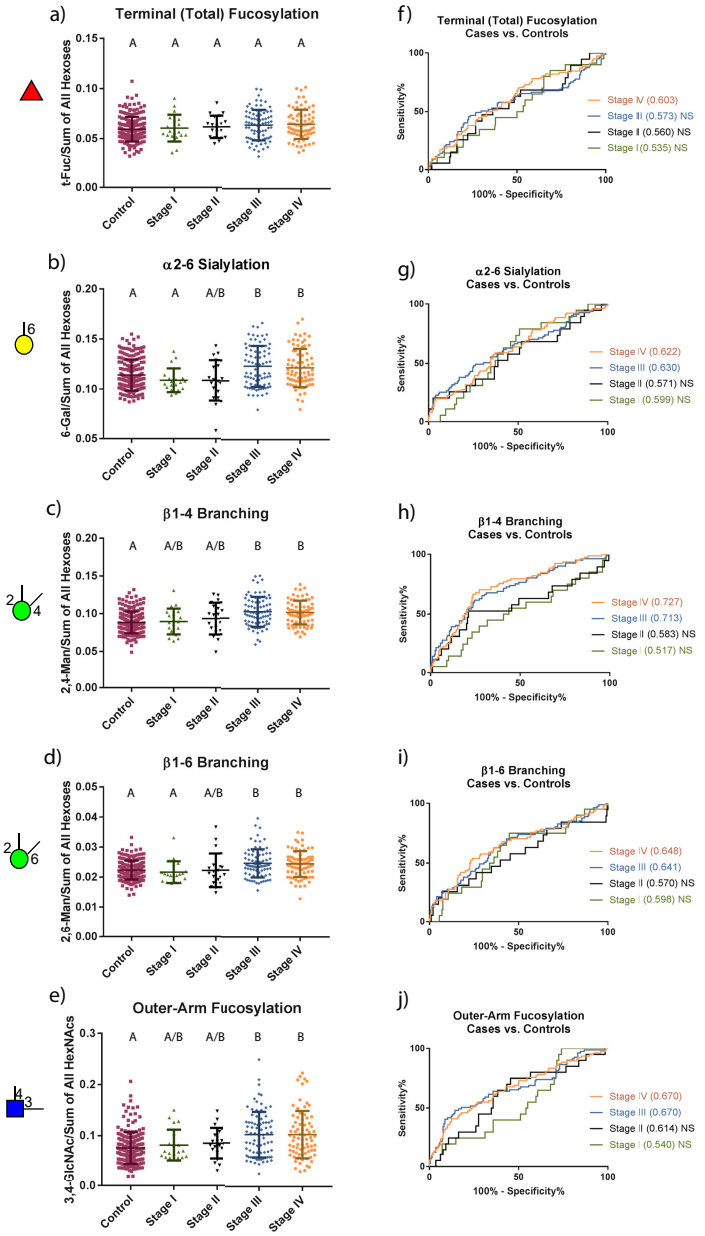

Figure 3.2. Univariate distributions and associated ROC curves for the top five-performing glycan nodes in the large lung cancer set when data were normalized to the sum of endogenous hexoses or HexNAcs. Letters above the data points in panels a-e indicate statistically significant differences between the five groups shown: any overlap in lettering between groups indicates a lack of significant difference between the groups (Kruskal-Wallis with Dunn's post hoc test). ROC curves for lung cancer cases (separated by stage) vs. controls are shown in panels f-j. ROC curve AUCs are provided in parenthesis next to the specified stages. "NS" next to ROC curve AUCs indicates that the ROC curve does not show a statistically significant difference between the two groups being compared. Glycan node symbol definitions are the same as in Figure 2.1.

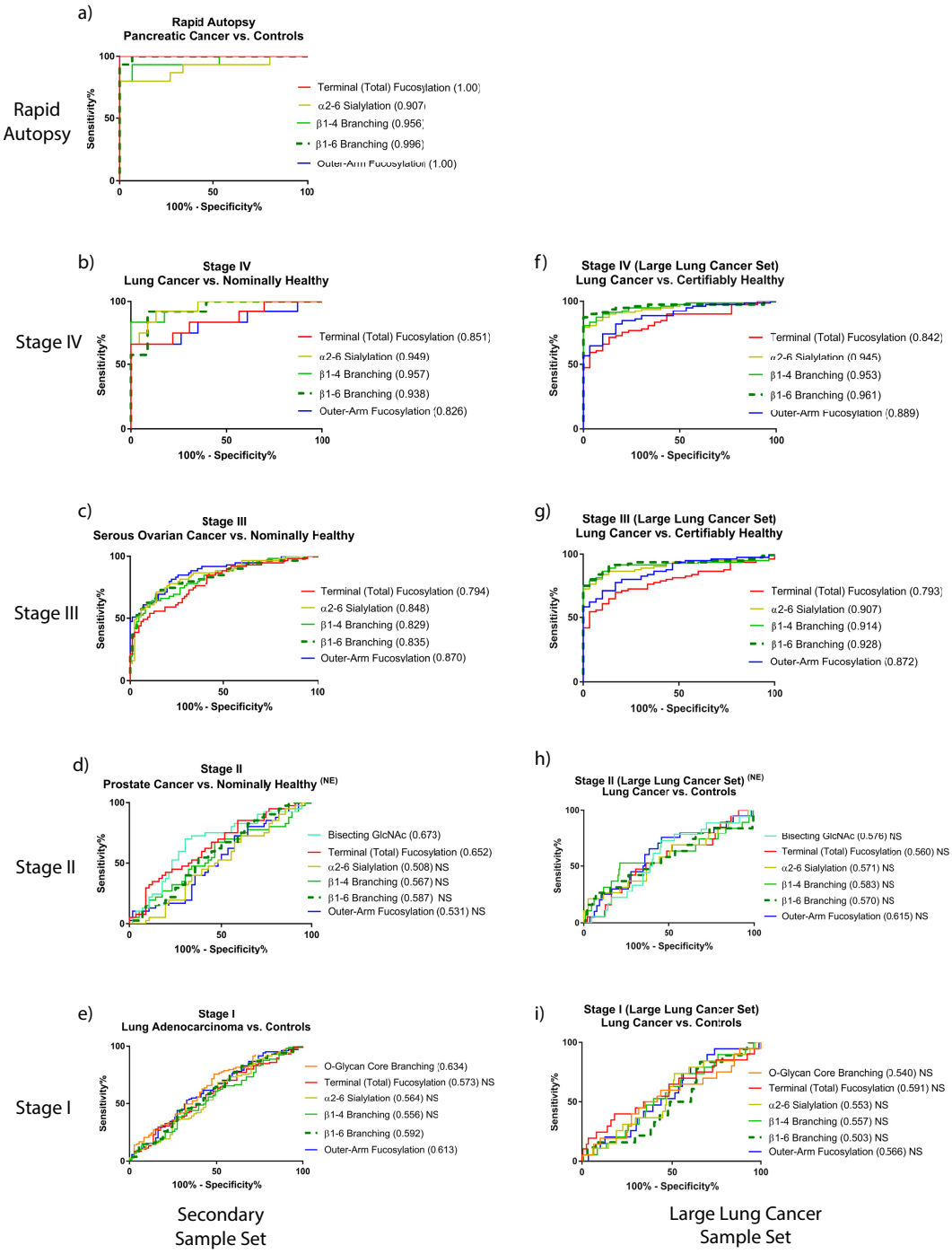
a)
Rapid Autopsy
Pancreatic Cancer vs. Controls
Terminal (Total) Fucosylation (1.00)
α2-6 Sialylation (0.907)
β1-4 Branching (0.956)
β1-6 Branching (0.996)
Outer-Arm Fucosylation (1.00)
Sensitivity%
100% - Specificity%
Rapid Autopsy
b)
Stage IV
Lung Cancer vs. Nominally Healthy
Terminal (Total) Fucosylation (0.851)
α2-6 Sialylation (0.949)
β1-4 Branching (0.957)
β1-6 Branching (0.938)
Outer-Arm Fucosylation (0.826)
Stage IV
c)
Stage III
Serous Ovarian Cancer vs. Nominally Healthy
Terminal (Total) Fucosylation (0.794)
α2-6 Sialylation (0.848)
β1-4 Branching (0.829)
β1-6 Branching (0.835)
Outer-Arm Fucosylation (0.870)
Stage III
d)
Stage II
Prostate Cancer vs. Nominally Healthy (NE)
Bisecting GlcNAc (0.673)
Terminal (Total) Fucosylation (0.652)
α2-6 Sialylation (0.508) NS
β1-4 Branching (0.567) NS
β1-6 Branching (0.587) NS
Outer-Arm Fucosylation (0.531) NS
Stage II
e)
Stage I
Lung Adenocarcinoma vs. Controls
O-Glycan Core Branching (0.634)
Terminal (Total) Fucosylation (0.573) NS
α2-6 Sialylation (0.564) NS
β1-4 Branching (0.556) NS
β1-6 Branching (0.592)
Outer-Arm Fucosylation (0.613)
Stage I
Secondary Sample Set
f)
Stage IV (Large Lung Cancer Set)
Lung Cancer vs. Certifiably Healthy
Terminal (Total) Fucosylation (0.842)
β1-4 Branching (0.953)
β1-6 Branching (0.961)
Outer-Arm Fucosylation (0.889)
g)
Stage III (Large Lung Cancer Set)
Lung Cancer vs. Certifiably Healthy
Terminal (Total) Fucosylation (0.793)
α2-6 Sialylation (0.907)
β1-4 Branching (0.914)
β1-6 Branching (0.928)
Outer-Arm Fucosylation (0.872)
h)
Stage II (Large Lung Cancer Set) (NE)
Lung Cancer vs. Controls
Terminal (Total) Fucosylation (0.560) NS
α2-6 Sialylation (0.571) NS
β1-4 Branching (0.583) NS
β1-6 Branching (0.570) NS
Outer-Arm Fucosylation (0.615) NS
i)
Stage I (Large Lung Cancer Set)
Lung Cancer vs. Controls
O-Glycan Core Branching (0.540) NS
Terminal (Total) Fucosylation (0.591) NS
α2-6 Sialylation (0.553) NS
β1-4 Branching (0.557) NS
β1-6 Branching (0.503) NS
Outer-Arm Fucosylation (0.566) NS
Large Lung Cancer Sample Set

Figure 3.3. ROC curves depicting the stage-dependent performance of the top five-performing glycan nodes in distinguishing different types of cancer from controls or healthy individuals (a-e). Adjacent to panels b-e are ROC curves from the large lung cancer study for comparison (f-i). Clear stage-dependence is evident, regardless of the type of cancer involved. A comparison of each ROC curve at each stage in the large lung cancer study to the parallel ROC curve in a different type of cancer or different lung cancer sample set revealed no significant differences between ROC curves (DeLong's test; see Table 2.8). A superscript "NE" (panels d and h) indicates that these data sets were normalized to the sum of endogenous hexoses or HexNAcs because heavy labeled internal standards were not added during analysis of the prostate cancer sample set.

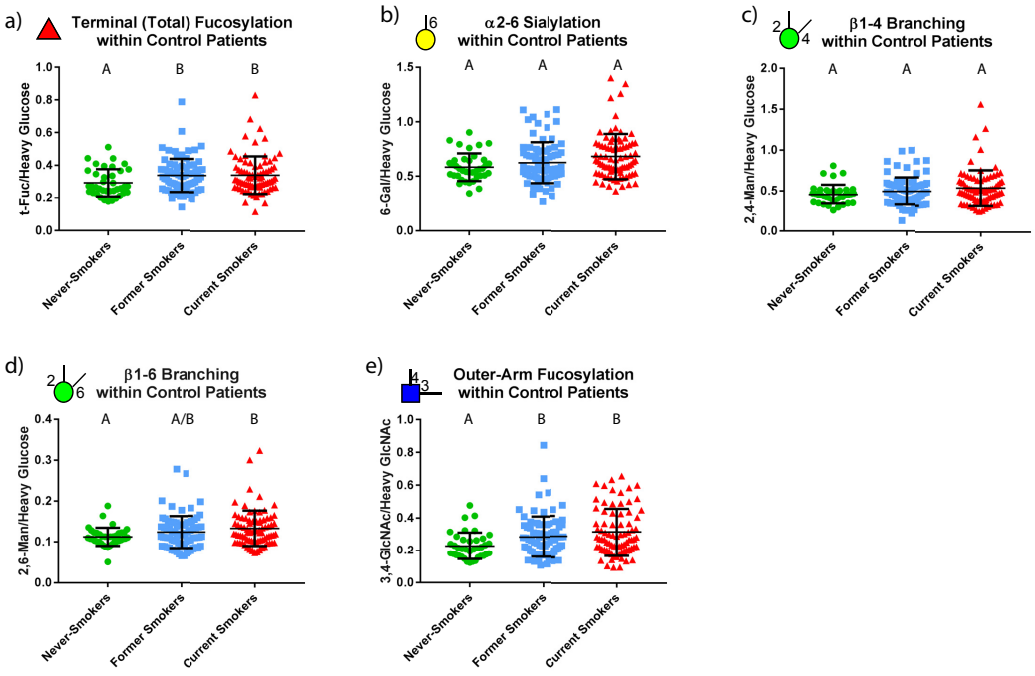

Figure 3.4. Univariate distributions of the top five-performing glycan nodes within the control group of the large lung cancer set, subdivided on the basis of smoking status. Letters above the data points indicate statistically significant differences between the three groups shown: any overlap in lettering between groups indicates a lack of significant difference between the groups (Kruskal-Wallis; Bonferroni-corrected p-values < 0.0167 for within-group pairwise comparisons were considered statistically significant).

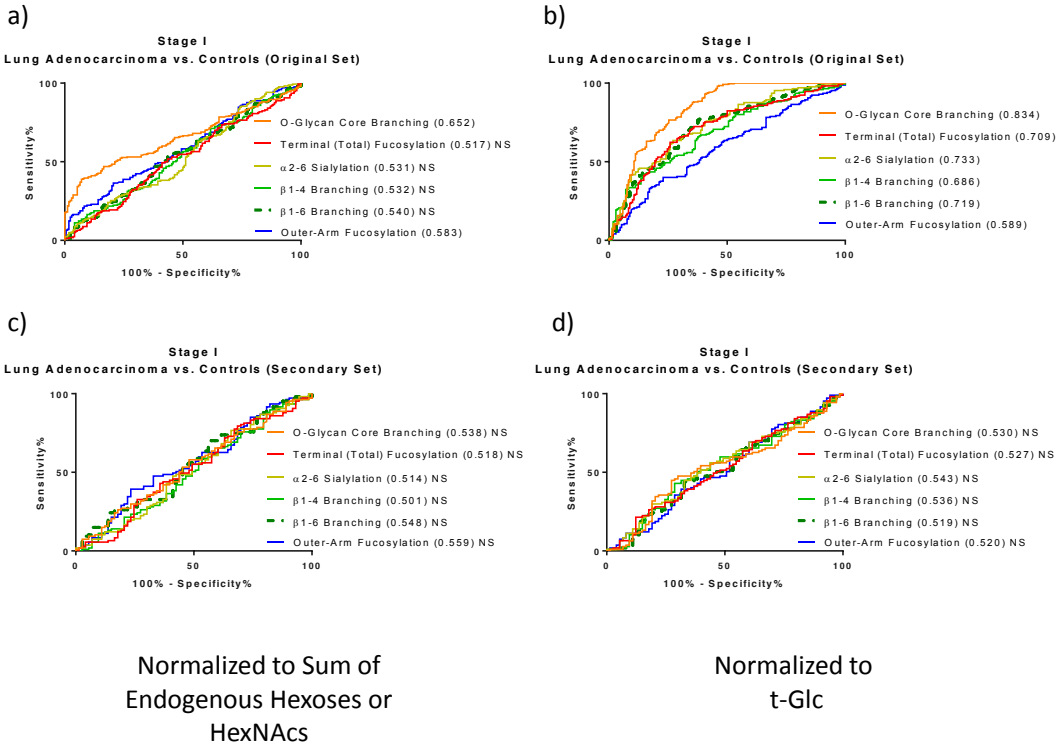

Figure 3.5. ROC curves illustrating the performance of top performing glycan nodes in distinguishing stage I lung adenocarcinoma from controls in the original set of samples (panels a and b) and the secondary set of samples (panels c and d). In panels a and c, data are normalized by sum of the hexoses or hexNAcs. In panels b and d data are normalized by t-Glc.

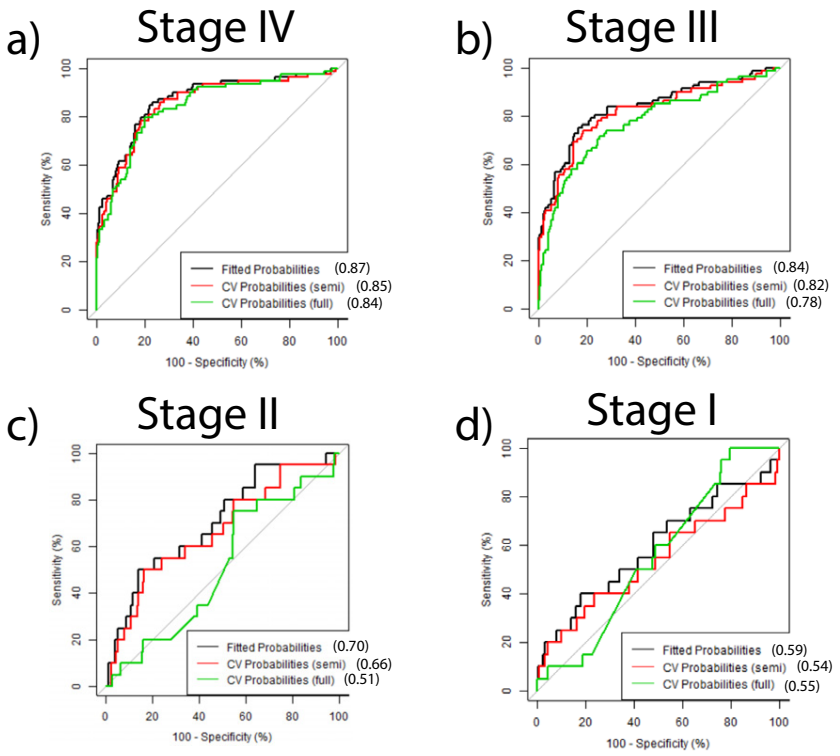

Figure 3.6. Multivariate logistic regression models for stage I-IV lung cancer patients from the large lung cancer data set. Fully validated multivariate combinations of glycan nodes did not produce significantly better ROC curves in stage IV, III, II, or I lung cancer patients (panels a-d, respectively) compared to the best performing individual glycan node in the control specimens (DeLong test). Three separate curves are shown on each plot, corresponding to predicted probabilities derived from a multivariate logistic regression model 1) re-fitted at each iteration of cross-validation (referred to as "CV Probabilities (full)"), 2) fitted once on the complete dataset, fixing the predictors, but allowing parameter estimates to change at each iteration of cross-validation (referred to as "CV Probabilities (semi)"), and 3) fitted once on the complete dataset and taking the model-derived probability without use of cross-validation (referred to as "Fitted Probabilities").

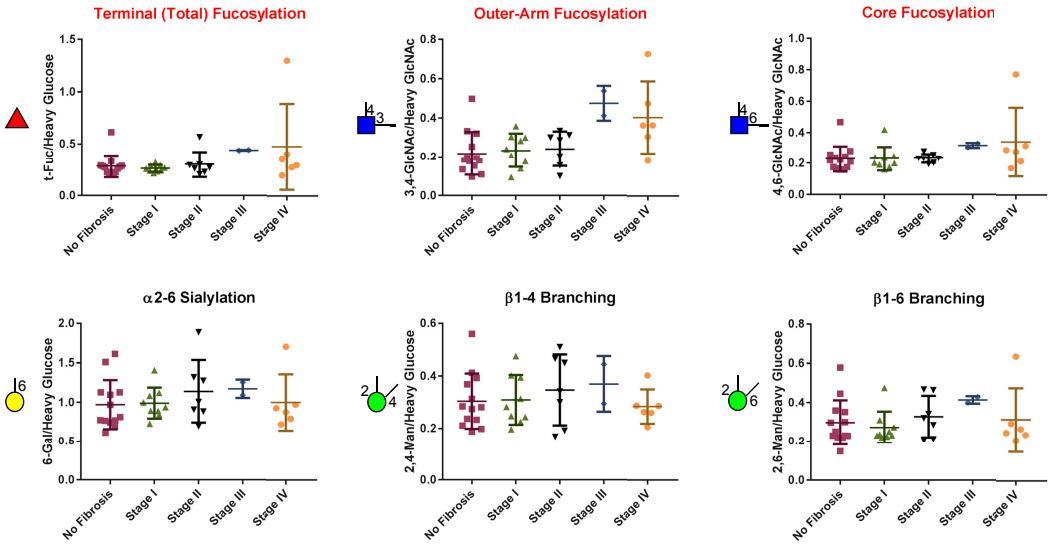

Figure 3.7. Univariate distributions of fucosylation-related glycan nodes, α2-6 sialylation, β1-4 branching and β1-6 branching in stage 0 through stage IV liver fibrosis. No statistically significant differences were observed for any pairwise comparisons within a single glycan node (Kruskal-Wallis).

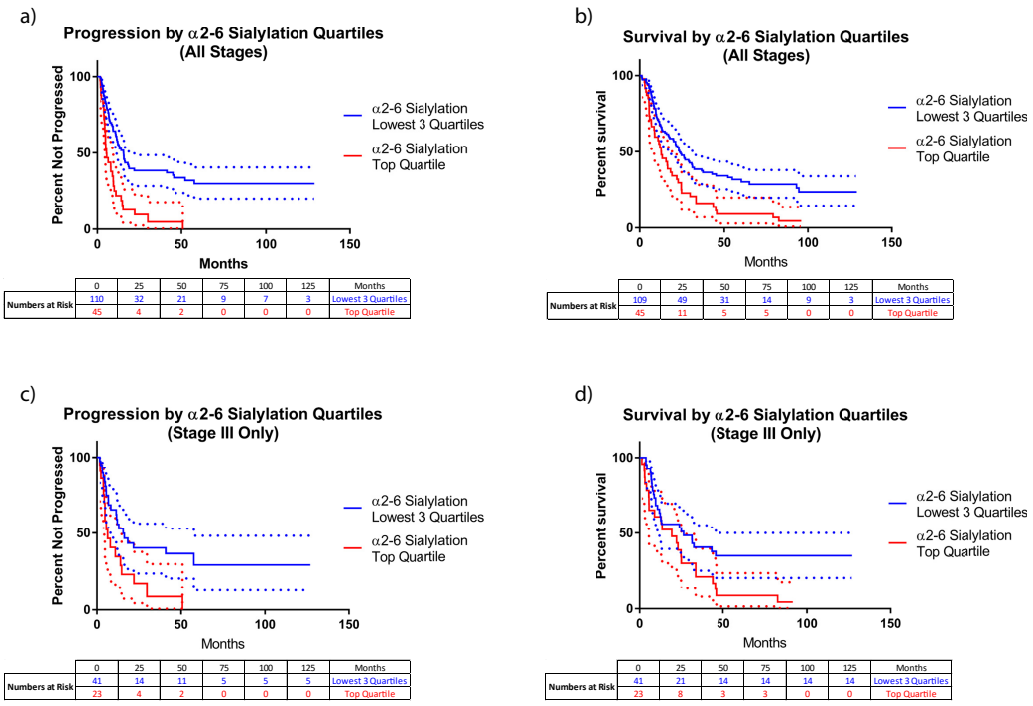

Figure 3.8. Large lung cancer data set progression (i.e., progression-free survival; a,c) and survival (all-cause mortality; b,d) curves for the top α2-6 sialylation quartile compared to all other quartiles combined. Panels a-b combine data from all stages; panels c-d present data from stage III only—illustrating that curve separation based on α2-6 sialylation is not simply driven by stage. Dotted lines represent 95% confidence intervals, colored according to their respective curves. Within each plot, progression curves were significantly different from one another (log-rank Mantel-Cox test; $p < 0.01$) as were the survival curves (log-rank Mantel-Cox test; $p < 0.05$). For the progression data (all stages; panel a), the median duration of follow-up for those that progressed, until progression, was 6.9 months (17.1 months median total follow-up time); for those that did not progress, the median duration of follow-up was 22.7 months. Results from Cox proportional hazards models are described in the Results section.

Table 3.1. Composition of sample sets and their sub-cohorts.

Name of Sample Set	Serum or Type of Plasma	Certifiably Healthy	Nominally Healthy [a]	Controls [b]	Stage I	Stage II	Stage III	Stage IV	Rapid Autopsy
Liver Kidney Donors	EDTA Plasma	30 [c]	-	-	-	-	-	-	-
Large Lung Cancer	Heparin Plasma [d]	-	-	199	20	20	81	78	-
Liver Fibrosis	Serum	-	-	13	9	7	2	6	-
Stage I Lung Adenocarcinoma (Original Set)	Serum	-	-	170	170	-	-	-	-
Stage I Adenocarcinoma (Secondary Set)	Serum	-	-	73	107	-	-	-	-
Stage II Prostate Cancer	Serum	-	56	-	-	40	-	-	-
Stage III Serous Ovarian Cancer	Serum	-	87	-	-	-	59	-	-
Stage IV Lung Cancer	Serum	-	23	-	-	-	-	12	-
Rapid Autopsy Pancreatic Cancer	Serum	-	-	15	-	-	-	-	15

[a] Nominally Healthy donors were self-reported as healthy and were only age and gender-matched to cases.
[b] Age & gender-matched to cases; smoking status-matched in lung cancer sets; benign nodule positive in Stage I Lung Adenocarcinoma set; benign inflammatory condition-matched in Pancreatic Cancer set
[c] These samples were analyzed immediately prior to—but in a separate set of batches from the Large Lung Cancer set. Results for the quality control specimens analyzed in both sets of batches were not significantly different from one another.
[d] Heparin is a glycosaminoglycan, but for reasons explained in the Methods section, it does not impact the results of glycan node analysis

Table 3.2. Basic clinical characteristics and n-values of the large lung cancer and certifiably healthy living kidney donors sample sets.

		Certifiably Healthy Living Kidney Donors[a]	Large Lung Cancer Set: Controls	Large Lung Cancer Set: Cases
Age [b]		46.5 ± 13.5[c]	60.5 ± 9.9	60.7 ± 10.3
Gender	Female	17	76	76
	Male	13	123	123
Smoking Status	Never-Smoker	22	38	34
	Former Smoker	8	80	70
	Current Smoker	0	81	95
Staging	Stage I	N/A	N/A	20
	Stage II	N/A	N/A	20
	Stage III	N/A	N/A	81
	Stage IV	N/A	N/A	78
Tumor Histology	Adenocarcinoma	N/A	N/A	79
	Squamous Cell Carcinoma	N/A	N/A	80
	Small Cell Carcinoma	N/A	N/A	40

[a] These specimens were collected as part of a separate study, see Methods section for additional details

[b] Age in years ± S.D.

[c] Significantly different from Controls and Cases (Kruskal-Wallis with Dunn's posthoc test; $p < 0.0001$)

Table 3.3. Statistically significant differences between cohorts within the large lung cancer study. Hexose data were normalized to heavy, stable isotope labeled glucose (Glc) and HexNAc data normalized to heavy, stable isotope labeled GlcNAc.

Glycan Node [a, b, c]	CH vs. C	CH vs. I	CH vs. II	CH vs. III	CH vs IV	C vs I	C vs. II	C vs. III	C vs. IV	I vs. II	I vs. III	I vs IV	II vs III	II vs IV	III vs IV
t-Fucose	ns	ii	ns	iiii	iiii	ii	ns	iiii	iiii	ns	ns	ns	ns	ns	ns
t-Gal	ns	ns	ns	i	iii	ns	ns	iii	iiii	ns	ns	ns	ns	ns	ns
2-Man	ns	ii	i	iiii	iiii	i	ns	iiii	iiii	ns	ns	ns	ns	ns	ns
4-Glc	dddd	dd	dddd	dddd	dddd	ii	ns	i	i	ns	ns	ns	ns	ns	ns
3-Gal	ns	ns	ns	i	i	ns	ns	i	i	ns	ns	ns	ns	ns	ns
6-Gal	i	ns	i	iiii	iiii	ns	ns	iiii	iiii	ns	iii	iii	ii	ii	ns
3,4-Gal	ns	ns	iiii	ns	ns	ns	iiii	i	i	iii	ns	ns	ii	ii	ns
2,4-Man	i	i	ii	iiii	iiii	ns	ns	iiii	iiii	ns	iii	iii	i	i	ns
2,6-Man	ii	i	iii	iiii	iiii	ns	ns	iiii	iiii	ns	iii	iii	i	i	ns
3,6-Man	ns	ns	ns	iii	iii	ns	ns	iiii	iiii	ns	ns	ns	ns	ns	ns
3,6-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,4,6-Man	i	iiii	i	ii	ii	ii	ns	ns	ns	ns	i	i	ns	ns	ns
t-GlcNAc	ns	ns	i	iiii	iiii	ns	i	iiii	iiii	ns	ii	i	ns	ns	ns
4-GlcNAc	i	ns	i	iiii	iiii	ns	ns	iiii	iiii	ns	ii	ii	ns	i	ns
3-GlcNAc	ns	ns	i	iiii	iiii	ns	ns	iiii	iiii	ns	iii	iii	ns	ns	ns
3-GalNAc	ii	ns	i	iiii	iiii	ns	ns	ii	iii	ns	ii	iii	ns	ii	ns
3,4-GlcNAc	ns	ns	i	iiii	iiii	ns	ns	iiii	iiii	ns	iii	iii	i	ii	ns
4,6-GlcNAc	ii	ns	ii	iiii	iiii	ns	ns	iii	iii	ns	ii	ii	ns	ns	ns
3,6-GalNAc	iiii	iii	iii	iiii	iiii	ns	ns	ii	iii	ns	ns	ns	ns	ns	ns

[a] Kruskal-Wallis test followed by Benjamini-Hochberg false discovery correction procedure where significance at the 95% confidence level is given by $p < 0.05$.
[b] "ns" indicates "not significant". "i" and "d" stand for "increased" or "decreased" in the clinically more-advanced cohort listed in the column header. i/d indicates $p < 0.05$; ii/dd indicates $p < 0.01$, iii/ddd indicates $p < 0.001$, iiii/dddd indicates $p < 0.0001$.
[c] CH: Certifiably Healthy, C: Controls, I: Stage I, II: Stage II, III: Stage III, IV: Stage IV

Table 3.4. Statistically significant differences between cohorts within the large lung cancer study. Hexose data were normalized to the sum of endogenous hexoses and HexNAc data were normalized to the sum of endogenous HexNAcs.

Glycan Node [a,b,c]	C vs I	C vs. II	C vs. III	C vs. IV	I vs. II	I vs. III	I vs IV	II vs III	II vs IV	III vs IV
t-Fucose	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
t-Gal	ns	ns	dddd	dddd	ns	ns	ns	ns	ns	ns
2-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4-Glc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	ii	ii	ns	ii	ii	ii	ii	ns
3,4-Gal	ns	i	ns	ns	ns	ns	ns	ns	ns	ns
2,4-Man	ns	ns	iiii	iiii	ns	ii	i	ns	ns	ns
2,6-Man	ns	ns	iii	iii	ns	i	i	i	ns	ns
3,6-Man	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,6-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3,4,6-Man	ns	ns	d	dd	ns	d	d	ns	d	ns
t-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	ddd	ddd	ns	ns	ns	ns	ns	ns
3,4-GlcNAc	ns	ns	iiii	iiii	ns	ns	ns	ns	ns	ns
4,6-GlcNAc	ns	ns	d	dd	ns	ns	ns	ns	ns	ns
3,6-GalNAc	ns	ns	dd	ddd	ns	ns	ns	ns	ns	ns

[a] Kruskal-Wallis test followed by Benjamini-Hochberg false discovery correction procedure where significance at the 95% confidence level is given by $p < 0.05$.
[b] "ns" indicates "not significant". "i" and "d" stand for "increased" or "decreased" in the clinically more-advanced cohort listed in the column header. i/d indicates $p < 0.05$; ii/dd indicates $p < 0.01$, iii/ddd indicates $p < 0.001$, iiii/dddd indicates $p < 0.0001$.
[c] CH: Certifiably Healthy, C: Controls, I: Stage I, II: Stage II, III: Stage III, IV: Stage IV

Table 3.5. Stage-by-stage comparison of the top performing glycan nodes. Comparisons are made for the large lung cancer set vs. other independent lung cancer sets (stages IV and I), ovarian (stage III) or prostate cancer (stage II). Actual ROC curves are shown in Fig. 2.5.

Stages	Glycan Feature	A: ROC AUC of Set A (Specified in Left Column) B: ROC AUC of Set B (Large Lung Cancer Set)	p-value of DeLong's test for two ROC curves [a]
Stage IV Sample Sets Compared (See Table 1): Set A: Stage IV Lung Cancer Set B: Large Lung Cancer	Terminal (Total) Fucosylation	A: 0.851 B: 0.841	0.907 (NS)
	α2-6 Sialylation	A: 0.949 B: 0.945	0.914 (NS)
	β1-4 Branching	A: 0.957 B: 0.953	0.937 (NS)
	β1-6 Branching	A: 0.938 B: 0.961	0.614 (NS)
	Outer-Arm Fucosylation	A: 0.826 B: 0.889	0.503 (NS)
Stage III Sample Sets Compared (See Table 1): Set A: Serous Ovarian Cancer Set B: Large Lung Cancer	Terminal (Total) Fucosylation	A: 0.794 B: 0.793	0.996 (NS)
	α2-6 Sialylation	A: 0.848 B: 0.907	0.179 (NS)
	β1-4 Branching	A: 0.829 B: 0.914	0.058 (NS)
	β1-6 Branching	A: 0.835 B: 0.928	0.035
	Outer-Arm Fucosylation	A: 0.870 B: 0.872	0.958 (NS)
Stage II Sample Sets Compared (See Table 1): Set A: Prostate Cancer Set B: Large Lung Cancer	Bisecting GlcNAc	A: 0.673 B: 0.576	0.253 (NS) [b]
	Terminal (Total) Fucosylation	A: 0.652 B: 0.560	0.298 (NS) [b]
	α2-6 Sialylation	A: 0.508 B: 0.571	0.507 (NS) [b]
	β1-4 Branching	A: 0.567 B: 0.583	0.872 (NS)
	β1-6 Branching	A: 0.587	0.867 (NS) [b]

[a] NS indicates no significant difference between ROC curves.
[b] *p*-value is the result of Bootstrap test since the Delong's test should not be applied to ROC curves with opposite directionalities

CHAPTER 4

ALTERATION OF SERUM GLYCAN FEATURES IN BLADDER CANCER

4.1 Introduction

Urothelial cell carcinoma (UCC) or bladder cancer is one of the top ten causes of cancer deaths annually [161]. From a clinical perspective, there are two major forms of this cancer: 1) non-muscle-invasive bladder cancer (NMIBC; stages pTa/pT1/pTis) and 2) muscle-invasive bladder cancer (MIBC; stages pT2+). Early detection of bladder cancer is very important; patients with non-muscle-invasive tumors have a much higher 5-year survival rate (88% for NMIBC patients relative to 41% for MIBC patients) [162]. Yet despite the stage at which it is diagnosed, high recurrence rate is one of the essential characteristics of this cancer [163]. Therefore, even if diagnosed at early stages and treated, former bladder cancer patients need to be monitored frequently. Currently, the common methods for detecting bladder cancer are cystoscopy (which is invasive and expensive[164]), urine cytology (which has low sensitivity for low-grade bladder cancer [165]), and computed tomography (CT) screening (which may not detect small tumors[166]). Accordingly, there has been a wide search for new biomarkers that are noninvasive, cost effective, and can outperform cytology [167]–[170]. At present there are no clinically employed serum-based markers for monitoring patients after their treatment. Plasma and serum (P/S) glycomics represents a promising source for a new generation of cancer markers [93]. Aberrant glycosylation is a universal feature of cancer [90] where it appears to enable the ability of tumor cells to avoid innate immune detection [171]. The changes in structure and abundance of glycans are often caused by dysregulated glycosyltransferase (GT) activity [49]. Conceptually, a glycan analysis technique that could provide one-to-one surrogate data for abnormal GT activity using routinely available clinical samples and that relied upon existing clinical technology

could be quite valuable.

As described above, in 2013, we developed a molecularly bottom-up approach that, unlike other approaches used in P/S glycomics, focuses on the analysis of monosaccharide and linkage-specific glycan "nodes" instead of intact glycans [88], [106], [107], [172]. It does this by employing the principles and processing chemistry of glycan methylation analysis (i.e., linkage analysis; **Figure 2.2**) to unfractionated P/S. This pools together each unique monosaccharide-and-linkage-specific glycan feature or glycan "node" from across all the normal and aberrant glycan structures in a given sample, providing a more direct surrogate measurement of GT activity than any single intact glycan. Moreover, many of these glycan nodes correspond directly and quantitatively to interesting glycan features such as "core fucosylation", "bisecting GlcNAc", and "β1-6 branching"—all captured as single GC-MS chromatographic peaks (**Figure 2.1**).

Our large lung cancer study provided important information about diagnostic and prognostic value of serum glycans in cancer [172]. To examine the potential clinical utility of this approach in the context of UCC, we applied glycan node analysis to 152 samples including 30 samples from certifiably healthy living kidney donors, 72 samples from patients with a history of bladder cancer but currently exhibiting no clinical evidence of disease (NED), 38 samples from patients with non-muscle-invasive bladder cancer, and 12 samples from patients with muscle invasive bladder cancer. Subsequently, we evaluated both the diagnostic and prognostic capacity of glycan "nodes" as clinical biomarkers. Elevated blood plasma protein glycosylation is known to be associated with inflammation in some non-cancerous clinical conditions [119]–[121]. Since C-reactive protein (CRP) is a well-studied marker of inflammation[173] as well as a prognostic

marker for UCC [174]–[177], we also evaluated the relationship between prognostically useful glycan nodes and CRP.

4.2 Materials and Method

4.2.1 Sample Information

EDTA plasma samples from certifiably healthy living kidney donors as well as current and former UCC patients were enrolled in the Multidisciplinary Biobank at Mayo Clinic Arizona under a Mayo Clinic IRB approved protocol. Patients eligible for enrollment were those seen at Mayo Clinic Arizona who were ≥ 18 years old, able to provide informed consent, and undergoing evaluation as either a potential living kidney donor or for genitourinary diseases. Detailed inclusion & exclusion criteria for living kidney donors are provided in Appendix A. None of the living kidney donor patients smoked at the time of health screening and blood collection; 27% were former smokers and 73% never smoked. Living kidney donor and UCC patients were excluded if they declined to participate or if the banking of their biospecimens would compromise the availability of tissue for diagnosis and standard clinical care. All specimens were collected during the time frame of June 2010 through Feb. 2016. Standard operating protocols and blood collections were performed as previously described [108]. All specimens were stored at -80 °C prior to shipment to ASU and maintained at -80 °C at ASU prior to analysis. All specimens were analyzed blind and in random order. An aliquot of plasma from the same individual donor was analyzed in every batch as a quality control (QC) specimen to ensure batch-to-batch consistency.

4.2.2 Glycan Node Analysis

Sample Preparation: Glycan node analysis was performed on the plasma samples as described previously (section 2.2.3). Briefly, it includes four main steps: 1)

permethylation, in which 9µL of plasma sample containing 1µL of a 10 mM solution of heavy-labeled D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-D_7, 97%–98%) (Cambridge Isotope Laboratories), and N-acetyl-D-[UL-$^{13}C_6$]glucosamine (Omicron Biochemicals, Inc.) as the internal standard was mixed with 270µL of dimethylsulfoxide (DMSO) (Sigma-Aldrich) followed by 105 µL of iodomethane (99%, Cat. No. I8507, Sigma-Aldrich). Then, this mixture was added to a plugged 1mL spin column (ThermoFisher Scientific, Waltham, MA, Cat. No. 69705) containing ~0.7g of sodium hydroxide beads (20–40 mesh, Sigma-Aldrich) which had been preconditioned by acetonitrile (Fisher Scientific) and washed twice with DMSO prior to addition of sample. After occasionally stirring the NaOH column over 11 min, the unplugged samples were spun for 15 s at 5,000 rpm (2,400g) in a microcentrifuge to extract the permethylated glycans. In order to maximize glycan recovery, 300µL of acetonitrile was added to the NaOH column and spun down for 30 s at 10,000 rpm (9,600g). Then, in a silanized 13 × 100 glass test tube holding 3.5 mL of 0.2 M sodium phosphate buffer, the solution from the first spin-through was added and mixed well. After pooling and mixing the second acetonitrile-based spin-through solution was combined with the rest of the sample, followed by 1.2 mL of chloroform (Sigma-Aldrich). The test tube was then capped and shaken well, followed by removal and discard of the aqueous layer. After two additional rounds of liquid/liquid extraction, the chloroform layer was recovered and dried under nitrogen at 74 °C. 2) The second major step was TFA hydrolysis, in which 325 µL of 2M trifluoroacetic acid (TFA) (Sigma-Aldrich) was added to each test tube. After capping the samples and incubating them at 121°C for 2h, they were dried down under nitrogen at 74 °C. 3) The third major step involved reduction of sugar aldehydes, in which the samples were incubated for an hour after adding 475 µL of a freshly made 10 mg/mL solution of sodium borohydride (Sigma-Aldrich) in 1M ammonium hydroxide (Sigma-Aldrich). Then 63 µL of methanol

(Honeywell Burdick & Jackson) was added to each sample before drying at 74 °C under nitrogen. Next, 125 µL of a 9:1 (v/v) methanol: acetic acid solution was added to each test tube followed by drying under nitrogen. To fully dry the samples, they were then placed in vacuum desiccator for approximately 20 min. 4) The fourth major step consisted of acetylation of nascent hydroxyl groups, in which the sample residue in each test tube was dissolved by 18 µL water before adding 250 µL of acetic anhydride (Sigma-Aldrich). After sonicating the samples for 2 min and incubating for 10 min at 60 °C, 230 µL of concentrated TFA was added to each sample, followed by incubation of the capped samples for 10 min at 60 °C. Then, 2 mL methylene chloride (Fisher Scientific) was added to each sample followed by 2 mL of water. Next, liquid/liquid extraction was done twice in which the methylene chloride layer was saved and then transferred into a silanized autosampler (ThermoFisher Scientific), dried under nitrogen, reconstituted in 120 µL of acetone (Avantor Performance Materials), and capped to be injected onto the GC-MS.

Gas Chromatography-Mass Spectrometry: An Agilent Model A7890 gas chromatograph (equipped with a CTC PAL autosampler) was used coupled to a Waters GCT (time-of-flight) mass spectrometer to analyze the prepared samples. For all samples, one injection of 1µL was made at split ratio of 20:1 onto an Agilent split-mode liner containing a small plug of silanized glass wool with the temperature set to 280 °C. The DB-5ms GC column that was used for chromatography was 30 m. The oven temperature, initially kept at 165 °C, was increased at a rate of 10 °C/min up to 265 °C. Immediately after that, the temperature was increased at a rate of 30 °C/min to 325 °C, then held constant for 3 min. The transfer line to the mass spectrometer was kept at 250 °C. Following the elution of sample components from the GC column, they were subjected to electron ionization (70 eV, 250 °C) and analyzed in the *m/z* range of 40-800

with a scan cycle time of 0.1 s. Daily calibration and tuning of the mass spectrometer was done using perfluorotributylamine.

The quantification method is described in detail elsewhere [88]. Briefly, summed extracted ion chromatogram peaks were integrated automatically and checked manually using QuanLynx software. The collected data were then exported to a spreadsheet for detailed analysis.

4.2.3 Human C-Reactive Protein ELISA Assay

The Invitrogen™ Human C-Reactive Protein ELISA kit (Catalog Number KHA0031, ThermoFisher Scientific) was used to measure the concentration of CRP in blood plasma samples from 30 certifiably healthy kidney donors, 72 patients with no evidence of disease, and 50 patients diagnosed with bladder cancer (38 NMIBC and 12 MIBC). The plasma samples were diluted 3000 times with standard diluent buffer. To prepare the standard solutions, Hu CRP standard was reconstituted to 8000 pg/mL using the standard diluent buffer. To make a 1200 pg/mL CRP solution, 300 µL of the reconstituted standard was added to a 2 mL eppendorf tube containing 1700 µL standard diluent buffer. Then, 300 µL of standard diluent buffer was added to each of the seven 1.5 mL eppendorf tubes labeled as 600, 300, 150, 75, 37.5, 18.75, and 0 pg/mL CRP to make serial dilutions of the standard, making the concentration of each solution as labeled on their tubes.

In order to perform ELISA, 100 µL of standards and pre-diluted samples were added to the appropriate wells of the antibody coated 96-well plate, leaving the wells for chromogen blanks empty. After covering the plate with a plate cover and incubating it for 2 h at 37 °C, the solution was thoroughly aspirated and the wells were washed with 1X wash buffer for 4 times. Then, 100 µL of Hu CRP Biotin Conjugate solution was added

into each well except the chromagen blanks. The plate was then covered and incubated for 1 h at room temperature, followed by aspiration of the solution and washing the plate. Before adding 100 µL 1X Streptavidin-HPR solution into each well except the chromogen blanks, the Streptavidin-HPR (100X) was diluted by adding 10 µL of it to a tube containing 1 mL Streptavidin-HPR Diluent for each 8-well strip used in the assay. Then, the covered plate was incubated for 30 min at room temperature. After aspirating the solution and washing the plate, 100 µL stabilized Chromogen was added to each plate, turning the substrate solution blue. Following 30 min incubation in the dark, 100 µL stop solution was added to each well. After mixing the solution by tapping the side of the plate, the solution in the wells changes from blue to yellow. The absorbance was read at 450 nm by Thermo Scientific Multiskan Go plate reader and the concentration of samples were calculated using SkanIt Software 3.2.

4.2.4 Statistical Analysis

Individual extracted-ion chromatographic peak areas for each glycan node were normalized using one of two possible approaches: 1) Individual hexose residues were normalized to heavy glucose and individual N-acetylhexosamine (HexNAc) residues were normalized to heavy N-acetyl glucosamine (heavy GlcNAc). 2) Individual hexose residues were normalized to the sum of all endogenous hexose residues. Likewise, each HexNAc residue was normalized to the sum of all endogenous HexNAcs. The average %CV calculated based on the analysis of the QC sample in each batch shows that the latter normalization method provides better within-batch reproducibility (< 10% for the top four performing glycan nodes) but the former normalization method performs better in separating the patient groups while still keeping the average inter-batch %CV in an

acceptable range (i.e., < 18%). Unless otherwise noted, results described below are based on normalization with heavy glucose and heavy GlcNAc.

For both the glycan node data and the CRP ELISA data, outliers within each clinical group (Control, NED, NMIBC and MIBC) were removed after $\log_{10}$ transformation using the ROUT method at Q = 1% by GraphPad Prism 7. After removing the outliers, the anti-log of each value was taken to reverse the transformation. To identify differences between clinical groups, the Kruskal-Wallis test was performed followed by the Benjamini-Hochberg false discovery correction procedure at a 5% false discovery rate using RStudio Version 1.0.143. Univariate distributions and ROC curves were plotted using GraphPad Prism 7. The ability of certain glycan nodes to predict bladder cancer recurrence was evaluated by performing Cox proportional hazards regression models using SAS 9.4. Correlations between CRP and glycan nodes were examined using Pearson correlation in GraphPad Prism 7.

4.3 Results

4.3.1 Altered Glycan Features in UCC

The relative abundance of 19 glycan "nodes" was quantified in each of the control, NED, NMIBC, and MIBC patient samples. Each of these nodes contributed at least 1% of the sum total of all hexoses or all HexNAcs. Data normalized to heavy, stable isotope-labeled glucose and GlcNAc internal standards were first evaluated for statistically significant differences between all four patient groups. Relative to the certifiably healthy controls, statistically significant changes were found in more than half of the glycan nodes measured in NED, NMIBC, and MIBC patients (**Table 4.1**). Among these glycan nodes, the only one that was *decreased* in the current and former cancer patient samples was 4-linked glucose (i.e., 4-Glc, which is mostly derived from glycolipids). The same trend was

previously observed in lung cancer patient samples [172]. The rest of the altered nodes were *increased* in current and former UCC patients compared to the certifiably healthy controls.

There were four glycan nodes that were most elevated in the current and former UCC patients relative to the certifiably healthy controls, including 6-linked galactose, 2,4-linked mannose, 2,6-linked mannose, and 3,4-linked GlcNAc. These nodes correspond to α2-6 sialylation, β1-4 branching, β1-6 branching, and outer-arm fucosylation, respectively [88], [172]. The univariate distributions of these four glycan nodes in each of the four clinical groups are shown in **Figure 4.1**. ROC curves for current and former UCC patients vs. the certifiably healthy controls are also shown (**Figure 4.1**). The distribution of each of these glycan nodes within each cohort shows that patients with no evidence of disease (NED) have similar glycosylation profiles to patients with active disease (NMIBC and MIBC) and that significant increases in these glycan nodes can only be seen when comparing current and former UCC patient to the certifiably healthy controls—but not when comparing amongst the three current and former UCC patient subgroups (**Table 4.1 and Figure 4.1**). Data normalized to the sum of endogenous hexoses or HexNAcs were not as effective at distinguishing the control specimens from those from current or former bladder cancer patients (**Table 4.2 and Figure 4.2**). However, as explained in the Discussion section, these data indicate that significant qualitative shifts in glycan composition are observed in current and former UCC patients as opposed to mere increases in the absolute abundance of glycans.

The average age of the certifiably healthy living kidney donors (controls) was 47, while the average age for the NED, NMIBC and MIBC patients was 74, 76, and 73, respectively. Yet, after correcting for multiple comparisons, no statistically significant

correlation of any glycan node with age could be found after pooling data from all cohorts and evaluating correlations for the age range in which there was overlap between the controls and the current and former UCC patients (i.e., ages 45-67; see **Figure 4.3**). Likewise, no significant correlations with age were observed within the certifiably healthy controls or within the current/former UCC patients when these groups were considered in isolation (not shown).

4.3.2 Prognostic Value of Glycan Nodes

Within the NED cohort there were numerous samples with high levels of specific glycan nodes that were well out of the range observed in the controls—and which were similar to the cancer patient samples—even though the NED patients were clinically free of disease (**Figure 4.1**). These observations led to evaluation of the ability of glycan nodes to predict recurrence in a Cox proportional hazards regression model. After breaking down glycan node data into quartiles and adjusting for age, gender, and time *from* cancer (i.e., time elapsed since there was evidence of cancer in a NED patient), 6-linked galactose and 2,6-linked mannose, which correspond to α2-6 sialylation and β1-6 branching, respectively, predicted recurrence with p-values of < 0.05. The top α2-6 sialylation quartile predicted recurrence from the NED state with a hazard ratio of 15 relative to all other quartiles combined (lower bound at 95% CL = 1.3; upper bound at 95% CL = 180; $p = 0.029$). Similarly, the top β1-6 branching quartile predicted recurrence from the NED state with a hazard ratio of 11 relative to all other quartiles combined (lower bound at 95% CL = 1.2; upper bound at 95% CL = 110; $p = 0.037$). The differences in the rates of recurrence for the top α2-6 sialylation and β1-6 branching quartiles compared to all other quartiles are shown in the progression-free survival curves (**Figure 4.4**).

4.3.3 CRP Correlation with Glycan Nodes

CRP was measured in order to correlate changes in patient glycan nodes with patient inflammation status. The average level of CRP in the certifiably healthy controls was 1.76 mg/L whereas the NED, NMIBC, and MIBC samples had average CRP levels of 3.84, 3.21, and 3.08 mg/L, respectively (which are above the normal range of CRP (<3.0 mg/L) [176]). The levels of 6-linked galactose, which corresponds to α2-6 sialylation, positively correlated with CRP ($r=0.34$, $p < 0.0001$), as did the levels of 2,6-linked mannose, which corresponds to β1-6 branching ($r=0.38$, $p <0.0001$) (**Figure 4.5**).

4.4 Discussion

Out of 19 quantified glycan nodes, four of them, each corresponding to a unique glycan feature including α2-6 sialylation, β1-4 branching, β1-6 branching and outer-arm fucosylation, were most significantly elevated in bladder cancer patients compared to certifiably healthy individuals (**Table 4.1 and Figure 4.1**). Interestingly, cancer-free patients with a history of UCC (NEDs) had glycan node distributions that were similar to both the early and late-stage cancer patients but distinct from the controls (**Figure 4.1**). And, unlike other types of cancers reported upon previously [172], glycan features were at the same level in later stages of UCC (MIBC patients) as in earlier stages (NMIBC patients).

In order to interpret these findings, it must be understood that the glycans being measured are from high-concentration glycoproteins derived primarily from the liver (i.e., transferrin, alpha-2-macroglobulin, haptoglobin, etc) and the immune system (i.e., IgG antibody glycans) rather than being sloughed off or secreted by cancer cells themselves [115], [116]. These macro-level (mg/mL scale) changes in blood plasma glycan biochemistry are thought to be mediated by cytokines secreted from the tumor

which are recognized by the liver and/or immune system as part of a systemic inflammatory response, altering the way that these two major glycoprotein-producing systems glycosylate their proteins [117], [133]–[137].

With this in mind, there are three possible causes for the increases in various glycan nodes observed in **Table 4.1** and **Figure 4.1**. First, the acute phase response in current and former UCC patients (evidenced by elevated CRP) may induce a net increase in the total concentration of plasma glycoproteins—and more glycoproteins means more glycans. Second, glycoprotein site occupancy may increase. While this possibility has not been extensively studied, some evidence exists that subtle but statistically significant increases in site occupancy may occur in steatosis and non-alcoholic steatohepatitis [178]. Third, the qualitative nature of the glycans themselves may change. This phenomenon has repeatedly been documented in cancer and is often the primary reason for shifts in glycan profiles—particularly when the data reported are compositional in nature (i.e., all signals sum to 100%) [120], [179], [180]. When the glycan node data from this study are normalized to the sum of endogenous hexose residues or HexNAc residues, statistically significant increases in the top four performing glycan nodes are observed in current and former UCC patients relative to the healthy controls (**Figure 4.2**)—though these increases tend not to be as strong as when total glycan node quantities are considered (**Figure 4.1**)—i.e., when the data are normalized to heavy labeled internal standards. Altogether, elevated CRP levels and the data seen in **Figure 4.2** suggest that both the first and the third possible explanations likely contribute to our observations. Assessing changes in glycan site occupancy requires establishing a complex, custom assay for each protein in question and is beyond the scope of the present study.

Overall, the glycan node distributions observed here in UCC suggest that UCC makes modest, early-stage alterations to blood plasma glycans that, even at stages III-IV, do not reach the extreme levels observed in pancreatic, lung, ovarian and other types of cancer [172] (Lung cancer patient glycan node data are compared side-by-side with UCC patient glycan node data in **Figure 4.6**). Yet even in remission, most former UCC patients with no evidence of disease (NED), tended to maintain these modestly elevated blood plasma glycan levels—wherein the NED patients with the most highly elevated levels were most likely to experience relapse (**Figure 4.4**).

It has previously been shown that serum glycans can be elevated in inflammatory patient states in the absence of cancer [119]–[121]. Moreover, chronic inflammation is known to be closely associated with the development of cancer [127], [128], [181]. Together with our observations, this suggests that the elevated plasma glycan levels observed in former UCC patients (currently in the NED state) that are prognostic of recurrence may be driven by or simply part of inflammatory processes. To assess this possibility, we measured CRP concentrations and found them to be strongly significantly correlated with levels of both α2-6 sialylation and β1-6 branching (**Figure 4.5**)—an observation that goes hand-in-hand with the fact that CRP has been found to predict UCC patient survival [176], [177].

This brings up the question of whether or not there is a mechanistic connection between alterations in plasma glycans (associated with inflammation) and the development or progression of cancer. There is evidence for the concept that the biological landscape experiences “grooming” or premetastatic “niche” formation prior to cancer establishing residence within the body [122]–[126]. And while glycans are not solely responsible for this process, evidence exists that they play important roles. As we have previously summarized [172] and others have explained in detail, cell-surface

glycans that facilitate resistance of galectin-mediated apoptosis [125], [138]–[140] (including poly-N-acetyllactosamine modified core 2 O-glycans [141]–[146]) as well as sialylated glycans that stimulate the inhibitory Siglec-7 receptor on natural killer cells [141]–[144] have important roles to play in helping cancer evade the body's natural immunity. Using glycan node analysis, we have observed major changes in plasma/serum glycans in lung, pancreatic, and ovarian cancers [172] as well as in UCC (this study). Further study is required to elucidate the potential mechanistic role of these macro-level changes to blood glycan biochemistry in the development and overall progression of cancer.

4.5 Conclusions

α2-6 sialylation, β1-4 branching, β1-6 branching, and outer-arm fucosylation were found to be significantly elevated in both current and former (in remission) UCC patients relative to certifiably healthy living kidney donors, with ROC curve c-statistics averaging approximately 0.8. In contrast to the stage-dependence that we have observed in other types of cancer [172], differences between patients with muscle invasive UCC, non-muscle invasive UCC and patients in remission were not statistically significant. For UCC patients in remission, α2-6 sialylation and β1-6 branching were prognostic indicators of recurrence and were correlated with CRP levels ($r = 0.34$ & 0.38, resp.; $p < 0.0001$), a known prognostic marker in UCC. Though glycan nodes exhibited less stage-dependency in UCC than in other cancers [172], results highlighted the pronounced difference between the serum glycan biochemistry of healthy individuals vs. any stage of UCC (including remission) and underscored the concept (previously observed [172]) that for plasma glycans the transition between a healthy state and an at-risk state is much more pronounced than that between an at-risk state and early stage cancer.

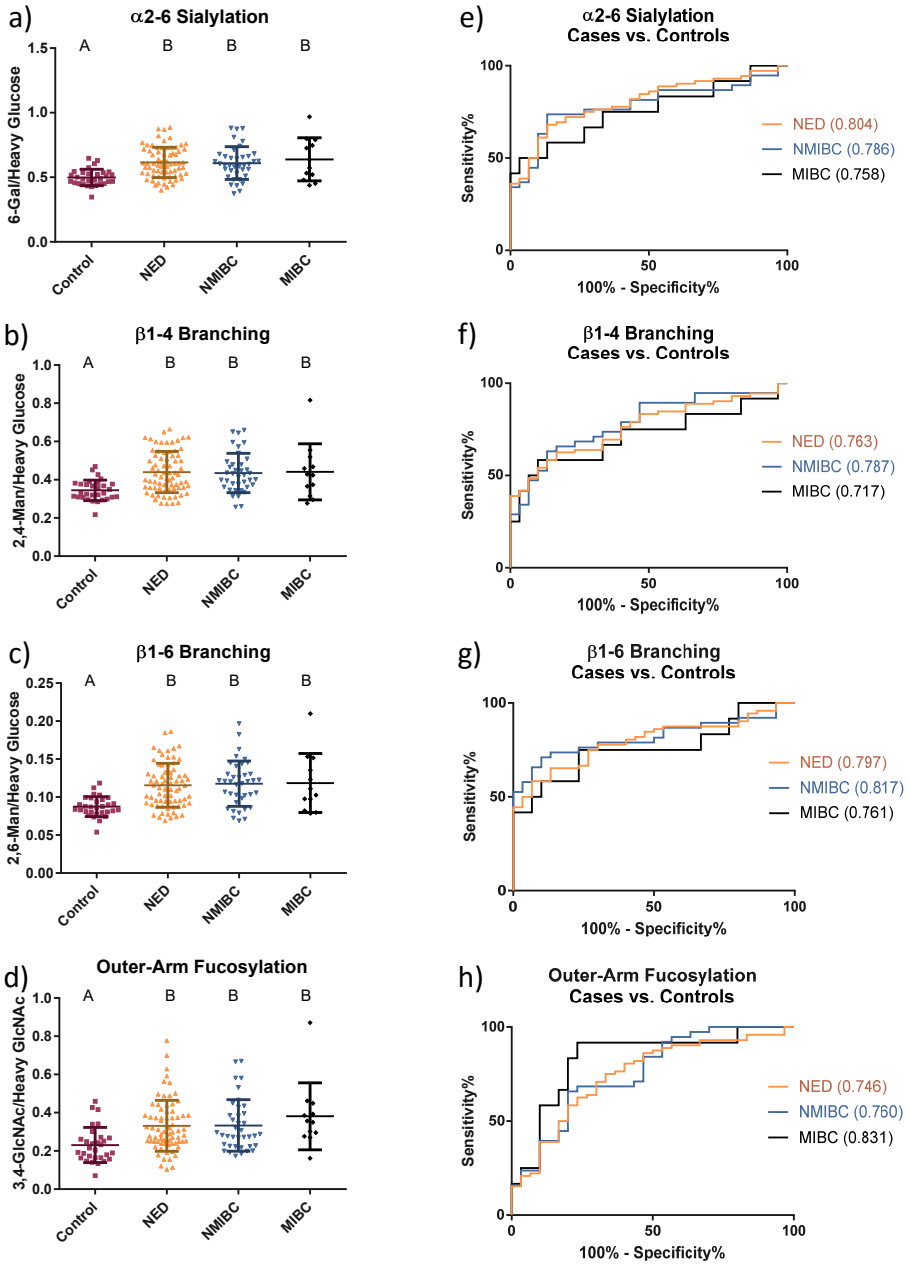

Figure 4.1. Distributions and ROC curves for the best performing (most diagnostic) glycan node markers in distinguishing different sub-cohorts of UCC patients from healthy controls when data were normalized to heavy glucose or heavy GlcNAc. Patient distributions are shown in (a-d). The Kruskal-Wallis test was performed followed by Dunn's post hoc test. The letters at the top of the data points show statistically significant differences between the patient groups; groups with same letter do not have a significant difference. (e-h) ROC curves for the different sub-cohorts of UCC patients vs. healthy individuals. Areas under the ROC curves are provided in parenthesis next to the stated patient groups.

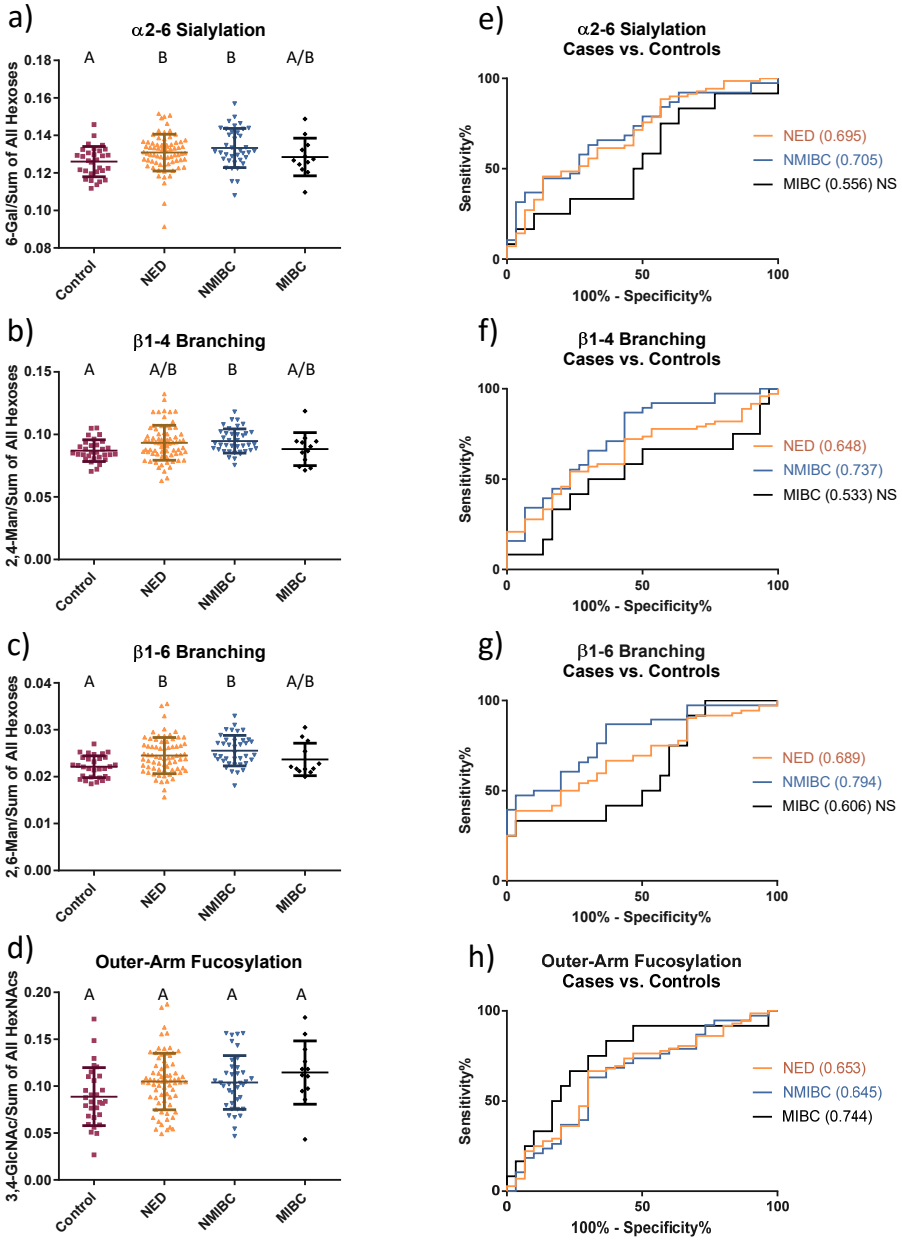

Figure 4.2. Distributions and ROC curves for the best performing glycan node markers in distinguishing different forms of bladder cancer from healthy controls when data were normalized to sum of endogenous Hexoses or HexNAcs. Patient distributions are shown in (a-d). The Kruskal-Wallis test was performed followed by Dunn's post hoc test. The letters at the top of the data points show statistically significant differences between the patient groups; groups with a common letter do not have a significant difference. (e-h) ROC curves for different groups of bladder cancer patients vs. certifiably healthy individuals. Area under the ROC curves are provided in parenthesis next to the stated patient groups. "NS" next to the area under the ROC curves shows that there is no significant difference between the two groups that are being compared.

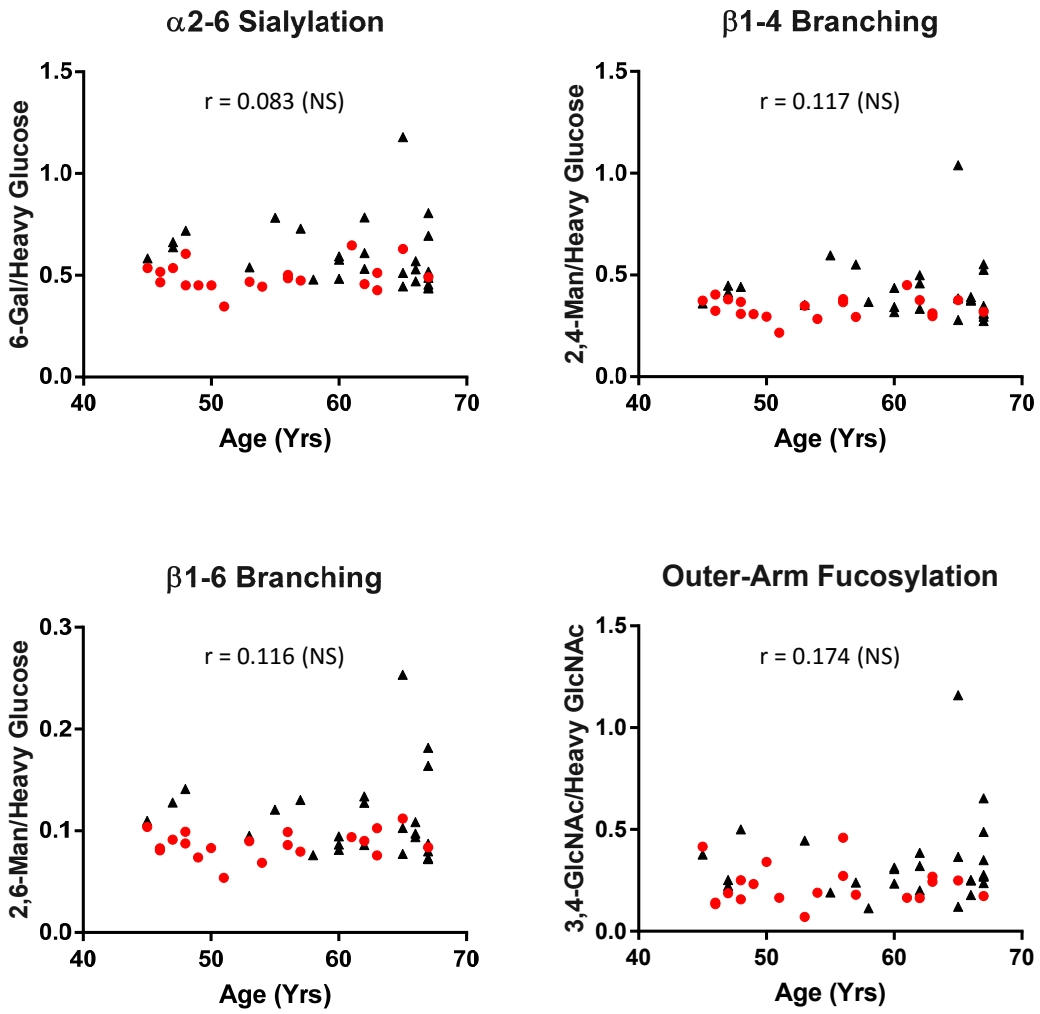

Figure 4.3. Correlation between age and the best performing (most diagnostic) glycan node markers when data were normalized to heavy glucose or heavy GlcNAc. Pearson correlation was used to evaluate this correlation. The common age range between all cohorts was 45-67. "NS" next to the r-value indicates that the Pearson correlation was not statistically significant. Distribution of the healthy controls is demonstrated by red dots. Distribution of the different sub-cohorts of UCC patients is demonstrated by black triangles.

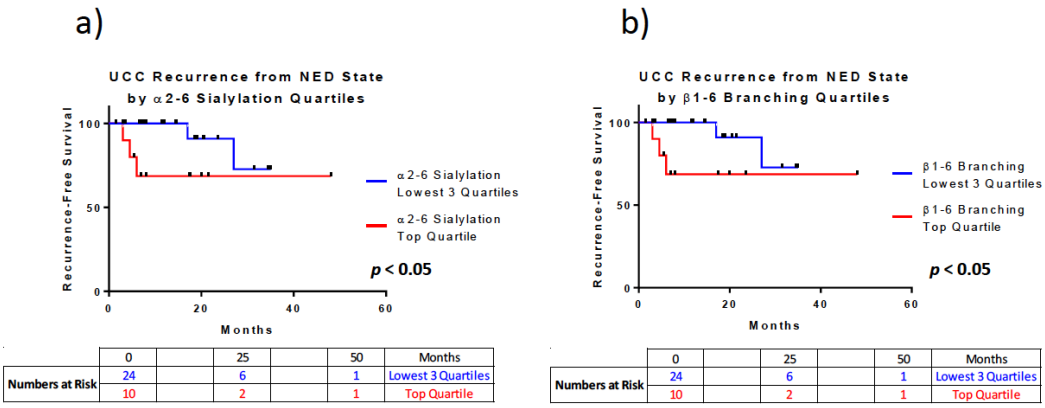

Figure 4.4. Bladder cancer recurrence curves for: (a) The top α2-6 sialylation quartile compared to all other quartiles combined. (b) The top β1-6 branching quartile compared to all other quartiles combined. In both panels, the recurrence curves within each plot were significantly different (log-rank Mantel-Cox test; $p < 0.05$). The median duration of follow-up for those that relapsed, until relapse, was 6 months, and for those that did not relapse was 12 months (The median total follow-up time was 11.75 months). The results of Cox proportional hazards models are reported in the Results section.

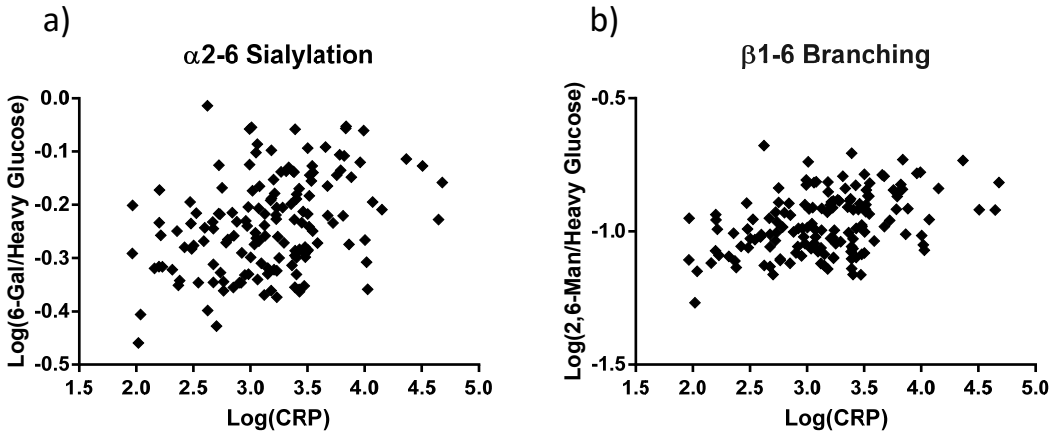

Figure 4.5. Correlation of CRP and glycan nodes. Log of CRP concentration vs. (a) α2-6 sialylation; r = 0.34 and (b) β1-6 branching; r =0.38 are plotted. Both correlations are statistically significant (Pearson correlation; $p < 0.01$).

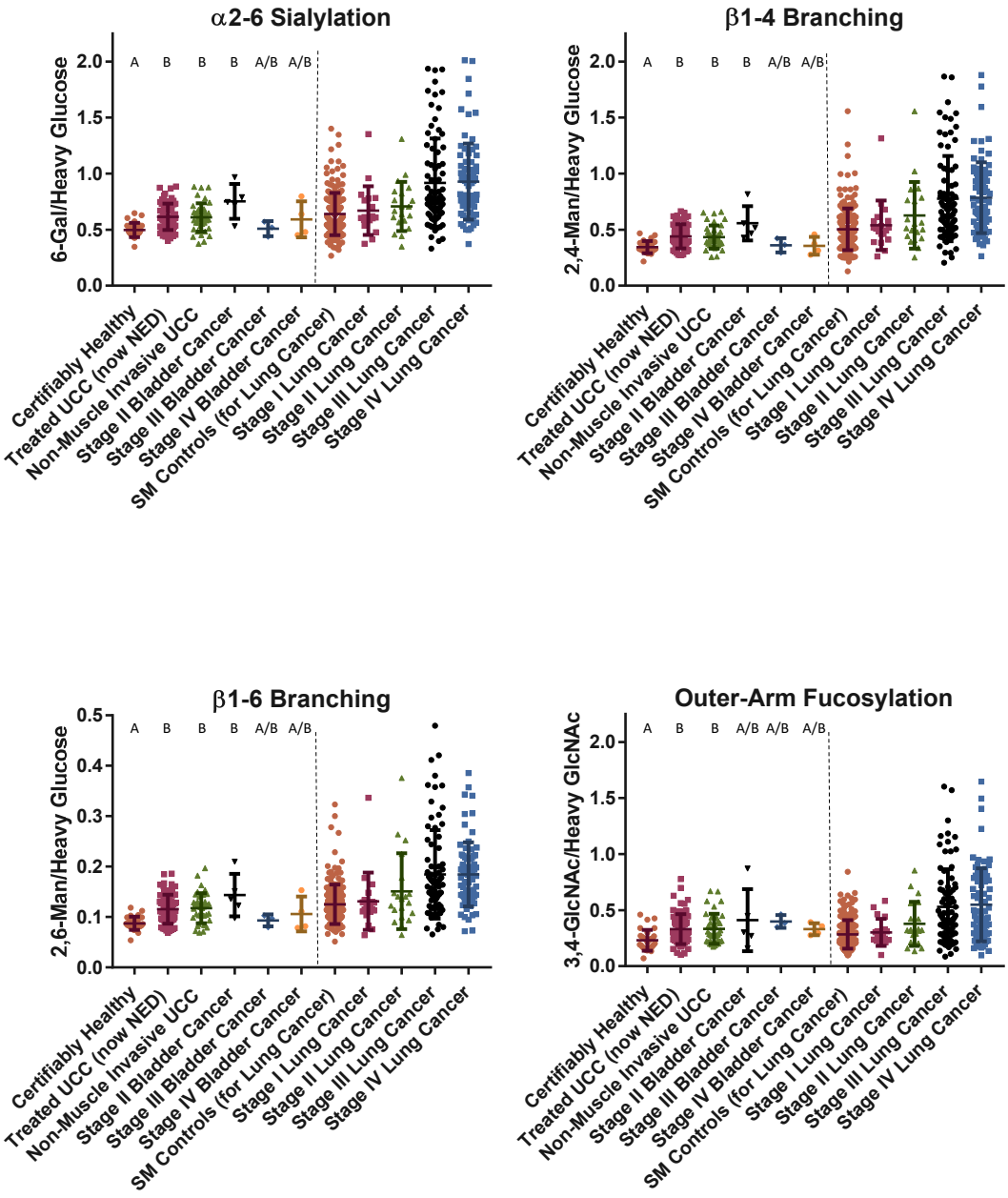

Figure 4.6. Distribution of the best performing glycan node markers in UCC with the MIBC group separated by patient stage. Data from a recently published lung cancer study [172] are displayed side-by-side for qualitative comparison. "SM Controls" indicates smoking status matched to the lung cancer patients. Letters at the top of each cohort show statistically significant differences between the patient groups; groups with a common letter do not have a significant difference.

Table 4.1. Statistically significant differences between controls and bladder cancer patient sub-cohorts. [a]

Glycan Nodes [b,c]	Control vs. NED	Control vs. NMIBC	Control vs. MIBC	NED vs. NMIBC	NED vs. MIBC	NMIBC vs MIBC
t-Fucose	i	i	ii	ns	ns	ns
t-Gal	ns	ns	ns	ns	ns	ns
2-Man	iii	ii	iii	ns	ns	ns
4-Glc	ns	d	ns	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns
6-Gal	iiii	iii	ii	ns	ns	ns
3,4-Gal	ns	ns	ns	ns	ns	ns
2,4-Man	iii	ii	i	ns	ns	ns
2,6-Man	iiii	iiii	ii	ns	ns	ns
3,6-Man	i	ns	ns	ns	ns	ns
3,6-Gal	ns	ns	ns	ns	ns	ns
3,4,6-Man	i	i	i	ns	ns	ns
t-GlcNAc	i	i	ns	ns	ns	ns
4-GlcNAc	ii	ii	i	ns	ns	ns
3-GlcNAc	ii	i	ii	ns	ns	ns
3-GalNAc	ns	ns	i	ns	ns	ns
3,4-GlcNAc	ii	ii	ii	ns	ns	ns
4,6-GlcNAc	ns	ns	ns	ns	ns	ns
3,6-GalNAc	iii	ii	iii	ns	ns	ns

[a] Individual hexose residues were normalized to heavy glucose and individual HexNAc residues were normalized to heavy GlcNAc)
[b] Significance was determined by the Kruskal-Wallis test followed by the Benjamini-Hochberg correction procedure at a 5% false discovery rate.
[c] "ns" stands for "not significant". "i" and "d" stand for "increased" or "decreased" glycan levels in the cohort with clinically more advanced disease listed in the column header. "i" or "d" indicates $p < 0.05$, "ii" or "dd" indicates $p < 0.01$, "iii" or "ddd" indicates $p < 0.001$, and "iiii" or "dddd" indicates $p < 0.0001$.

Table 4.2. Statistically significant differences between controls and bladder cancer patient sub-cohorts with data normalization to the sum of all endogenous hexoses or HexNAcs.

Glycan Nodes [a,b]	Control vs. NED	Control vs. NMIBC	Control vs. MIBC	NED vs. NMIBC	NED vs. MIBC	NMIBC vs MIBC
t-Fucose	ns	ns	ns	ns	ns	ns
t-Gal	dd	dd	ns	ns	ns	ns
2-Man	ns	ns	ns	ns	ns	ns
4-Glc	ddd	dddd	d	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns
6-Gal	i	ii	ns	ns	ns	ns
3,4-Gal	ns	ns	ns	ns	ns	ns
2,4-Man	ns	ns	ns	ns	ns	ns
2,6-Man	ii	iii	ns	ns	ns	ns
3,6-Man	ns	ns	ns	ns	ns	ns
3,6-Gal	ns	ns	ns	ns	ns	ns
3,4,6-Man	ns	ns	ns	ns	ns	ns
t-GlcNAc	ns	ns	ns	ns	ns	ns
4-GlcNAc	ns	ns	ns	ns	ns	ns
3-GlcNAc	ns	ns	ns	ns	ns	ns
3-GalNAc	dd	d	ns	ns	ns	ns
3,4-GlcNAc	i	ns	i	ns	ns	ns
4,6-GlcNAc	ns	dd	ns	ns	ns	ns
3,6-GalNAc	ns	ns	ns	ns	ns	ns

[a] Significance was determined by the Kruskal-Wallis test followed by the Benjamini-Hochberg correction procedure at a 5% false discovery rate.
[b] "ns" stands for "not significant". "i" and "d" stand for "increased" or "decreased" glycan levels in the cohort with clinically more advanced disease listed in the column header. "i" or "d" indicates $p < 0.05$, "ii" or "dd" indicates $p < 0.01$, "iii" or "ddd" indicates $p < 0.001$, and "iiii" or "dddd" indicates $p < 0.0001$.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, we assessed the behavior of P/S glycans in different types of cancer across various stages by applying our molecularly bottom-up P/S glycomics approach known as "glycan node analysis" to over a thousand clinical P/S samples from different case control studies across all stages of cancer. Before initiating the studies on cancer sample sets, in order to validate the robustness of our method, glycan nodes were analyzed in a matched collection of serum and several different types of plasma to determine whether subtle differences in sample matrix impacted the analytical results. Sodium citrate and sodium EDTA plasma samples were excluded from the studies since their results were not consistent with serum and other types of plasma samples. Furthermore, analysis of glycan nodes from samples treated in different conditions relative to controls that were stored at -80 °C the whole time, suggested that glycans are fairly stable in plasma and serum.

In lung cancer, the top performing glycan nodes in distinguishing cases from controls were including terminal (total) fucosylation, α2–6 sialylation, β1–4 branching, β1–6 branching, and outer-arm fucosylation. The elevation of these markers were stage dependent, yet independent of the tumor tissue-of-origin (when compared to ovarian, prostate, and pancreatic cancer). Out of these top performing glycan features, α2–6 sialylation was able to predict both progression and all-cause mortality in lung cancer patients using a Cox proportional hazards regression model adjusted for age, gender, smoking status, and cancer stage. Moreover, markers such as α2–6 sialylation, β1–4 branching, β1–6 branching were significantly increased in risk-matched controls relative to certifiably healthy living kidney donors.

In bladder cancer, markers for α2-6 sialylation, β1-4 branching, β1-6 branching, and outer-arm fucosylation were able to separate current and former (NED) cases from certifiably healthy controls; but since patients with a history of UCC had similar glycan distributions to both early and late stage cancer patients (which unlike other types of cancers mentioned earlier, had the same level of glycans), NED, NMIBC, and MIBC were not distinguished from one another. Furthermore, Using a Cox proportional hazards regression model adjusted for age, gender, and time from cancer, α2-6 sialylation and β1-4 branching were found to predict recurrence from the NED state and were correlated with CRP levels.

Comparing our findings from the lung cancer study to the bladder cancer study, we realized that the glycan node levels in UCC do not alter in a stage-dependent manner; unlike lung and other types of cancers involved in the lung cancer study. They modestly alter in early stage UCC and even at stages III-IV do not reach the extreme levels observed in other types of cancer. The reason for this behavior in P/S glycans of UCC patients is not fully understood and needs to be further studied.

Our observations together with the information provided by other studies [119]–[121], [127], [128], [181] suggest that the elevated levels of many glycan features in the at-risk controls (lung cancer study) as well as former UCC patients (NEDs) is driven by or simply part of inflammatory processes. These differences in the at-risk controls and former UCC patients compared to healthy individuals support the idea that the biological landscape undergoes “grooming” or premetastatic “niche” formation prior to cancer development within the body [122]–[126]. Further study is required to elucidate the potential mechanistic role of these changes to blood glycan biochemistry in the development and overall progression of cancer.

REFERENCES

[1] A. Varki, “Biological roles of glycans,” *Glycobiology*, vol. 27, no. 1, pp. 3–49, 2017.

[2] R. G. Spiro, “Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds,” *Glycobiology*, vol. 12, no. 4, p. 43R–56R, 2002.

[3] D. J. Moloney *et al.*, “Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules,” *J. Biol. Chem.*, vol. 275, no. 13, pp. 9604–9611, 2000.

[4] J. A. Goetz, M. V. Novotny, and Y. Mechref, “Enzymatic/chemical release of O-glycans allowing MS analysis at high sensitivity,” *Anal. Chem.*, vol. 81, no. 23, pp. 9546–9552, 2009.

[5] F. Maley, R. B. Trimble, A. L. Tarentino, and T. H. Plummer, “Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases,” *Analytical Biochemistry*, vol. 180, no. 2. pp. 195–204, 1989.

[6] X. Song, H. Ju, Y. Lasanajak, M. R. Kudelka, D. F. Smith, and R. D. Cummings, “Oxidative release of natural glycans for functional glycomics,” *Nat. Methods*, vol. 13, no. 6, pp. 528–534, 2016.

[7] H. E. McFarlane, A. Döring, and S. Persson, “The Cell Biology of Cellulose Synthesis,” *Annu. Rev. Plant Biol.*, vol. 65, no. 1, pp. 69–94, 2014.

[8] B. E. Koch, J. Stougaard, and H. P. Spaink, “Keeping track of the growing number of biological functions of chitin and its interaction partners in biomedical research,” *Glycobiology*, vol. 25, no. 5. pp. 469–482, 2015.

[9] M. J. Mitchell and M. R. King, “Physical biology in cancer. 3. The role of cell glycocalyx in vascular transport of circulating tumor cells.,” *Am. J. Physiol. Cell Physiol.*, vol. 306, no. 2, pp. C89-97, 2014.

[10] A. Varki and J. B. Lowe, “Chapter 6. Biological Roles of Glycans,” in *Essentials of Glycobiology*, 2009.

[11] R. Leavitt, S. Schlesinger, and S. Kornfeld, “Impaired intracellular migration and altered solubility of nonglycosylated glycoproteins of vesicular stomatitis virus and Sindbis virus,” *J. Biol. Chem.*, vol. 252, no. 24, pp. 9018–9023, 1977.

[12] A. Rapraeger, A. Krufka, and B. Olwin, “Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation,” *Science (80-.).*, vol. 252, no. 5013, pp. 1705–1708, 1991.

[13] X. Wang *et al.*, “From The Cover: Dysregulation of TGF- 1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice,” *Proc. Natl. Acad. Sci.*, vol. 102, no. 44, pp. 15791–15796, 2005.

[14] F. Y. Avci, X. Li, M. Tsuji, and D. L. Kasper, “Carbohydrates and T cells: A sweet twosome,” *Seminars in Immunology*, vol. 25, no. 2. pp. 146–151, 2013.

[15] C.-C. Wang *et al.*, “Glycans on influenza hemagglutinin affect receptor binding and immune response,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 43, pp. 18137–18142, 2009.

[16] M. Demetriou, M. Granovsky, S. Quaggin, and J. W. Dennis, “Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation,” *Nature*, vol. 409, no. 23, pp. 3789–3794, 2001.

[17] J. W. Dennis, S. Laferté, C. Waghorne, M. L. Breitman, and R. S. Kerbel, “Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis.,” *Science*, vol. 236, no. 4801, pp. 582–585, 1987.

[18] Y. Song, J. A. Aglipay, J. D. Bernstein, S. Goswami, and P. Stanley, “The bisecting GlcNAc on N-glycans inhibits growth factor signaling and retards mammary tumor progression,” *Cancer Res.*, vol. 70, no. 8, pp. 3361–3371, 2010.

[19] R. Singhai, V. W. Patil, S. R. Jaiswal, S. D. Patil, M. B. Tayade, and A. V. Patil, “E-Cadherin as a diagnostic biomarker in breast cancer,” *N. Am. J. Med. Sci.*, vol. 3, no. 5, pp. 227–233, 2011.

[20] J. W. Dennis, I. R. Nabi, and M. Demetriou, “Metabolism, Cell Surface Organization, and Disease,” *Cell*, vol. 139, no. 7. pp. 1229–1241, 2009.

[21] K. S. Lau *et al.*, “Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation,” *Cell*, vol. 129, no. 1, pp. 123–134, 2007.

[22] D. Ilver, “Helicobacter pylori Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging,” *Science (80-.).*, vol. 279, no. 5349, pp. 373–377, 1998.

[23] X. Jiang *et al.*, “Antiadhesion therapy for urinary tract infections-A balanced PK/PD profile proved to be key for success,” *J. Med. Chem.*, vol. 55, no. 10, pp. 4700–4713, 2012.

[24] P. A. Orlandi, F. W. Klotz, and J. D. Haynes, “A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac(α2-3)Gal- sequences of glycophorin A,” *J. Cell Biol.*, vol. 116, no. 4, pp. 901–909, 1992.

[25] M. L. Zupancic, M. Frieman, D. Smith, R. A. Alvarez, R. D. Cummings, and B. P. Cormack, “Glycan microarray analysis of Candida glabrata adhesin ligand specificity,” *Mol. Microbiol.*, vol. 68, no. 3, pp. 547–559, 2008.

[26] S. K. Mazmanian, H. L. Cui, A. O. Tzianabos, and D. L. Kasper, “An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system,” *Cell*, vol. 122, no. 1, pp. 107–118, 2005.

[27] I. van Die and R. D. Cummings, “Glycan gimmickry by parasitic helminths: A strategy for modulating the host immune response?,” *Glycobiology*, vol. 20, no. 1. pp. 2–12, 2010.

[28] S. J. van Vliet *et al.*, “Carbohydrate profiling reveals a distinctive role for the C-type lectin MGL in the recognition of helminth parasites and tumor antigens by dendritic cells,” *Int. Immunol.*, vol. 17, no. 5, pp. 661–669, 2005.

[29] T. Kawano, “CD1d-Restricted and TCR-Mediated Activation of V14 NKT Cells by Glycosylceramides,” *Science (80-.).*, vol. 278, no. 5343, pp. 1626–1629, 1997.

[30] A. Varki and S. Kornfeld, “P-type Lectins,” *Essentials Glycobiol. 2nd ed.*, pp. 425–437, 2009.

[31] E. Li, I. Tabas, and S. Kornfeld, “The synthesis of complex-type oligosaccharides. I. Structure of the lipid-linked oligosaccharide precursor of the complex-type oligosaccharides of the vesicular stomatitis virus G protein.,” *J. Biol. Chem.*, vol. 253, no. 21, pp. 7762–7770, 1978.

[32] A. J. Parodi, D. H. Mendelzon, and G. Z. Lederkremer, “Transient glucosylation of protein-bound Man9GlcNAc2, Man8GlcNAc2, and Man7GlcNAc2 in calf thyroid cells. A possible recognition signal in the processing of glycoproteins,” *J. Biol. Chem.*, vol. 258, no. 13, pp. 8260–8265, 1983.

[33] C. Hammond, I. Braakman, and A. Helenius, “Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control.,” *Proc. Natl. Acad. Sci.*, vol. 91, no. 3, pp. 913–917, 1994.

[34] A. C. Castonguay, L. J. Olson, and N. M. Dahms, “Mannose 6-phosphate receptor homology (MRH) domain-containing lectins in the secretory pathway,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1810, no. 9. pp. 815–826, 2011.

[35] B. Zhang *et al.*, “Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex,” *Nat Genet*, vol. 34, no. 2, pp. 220–225, 2003.

[36] M. G. Wahrenbrock and A. Varki, “Multiple hepatic receptors cooperate to eliminate secretory mucins aberrantly entering the bloodstream: Are circulating cancer mucins the ‘tip of the iceberg’?,” *Cancer Res.*, vol. 66, no. 4, pp. 2433–2441, 2006.

[37] J. D. Murray, “Invasion by invitation: Rhizobial infection in legumes,” *Mol. Plant-Microbe Interact.*, vol. 24, no. 6, pp. 631–639, 2011.

[38] T. Marquardt, K. Lühn, G. Srikrishna, H. H. Freeze, E. Harms, and D. Vestweber, “Correction of leukocyte adhesion deficiency type II with oral fucose.,” *Blood*, vol. 94, no. 12, pp. 3976–85, 1999.

[39] M. J. Telen *et al.*, “Randomized phase 2 study of GMI-1070 in SCD: Reduction in time to resolution of vaso-occlusive events and decreased opioid use,” *Blood*, vol. 125, no. 17, pp. 2656–2664, 2015.

[40] A. F. Carlin, S. Uchiyama, Y. C. Chang, A. L. Lewis, V. Nizet, and A. Varki, “Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response,” *Blood*, vol. 113, no. 14, pp. 3333–3336, 2009.

[41] B. Khatua, K. Bhattacharya, and C. Mandal, “Sialoglycoproteins adsorbed by Pseudomonas aeruginosa facilitate their survival by impeding neutrophil extracellular trap through siglec-9,” *J. Leukoc. Biol.*, vol. 91, no. 4, pp. 641–655, 2012.

[42] J. W. Johnston *et al.*, “Regulation of sialic acid transport and catabolism in Haemophilus influenzae,” *Mol. Microbiol.*, vol. 66, no. 1, pp. 26–39, 2007.

[43] L. Freire-De-Lima, L. M. Fonseca, T. Oeltmann, L. Mendonçą-Previato, and J. O. Previato, “The trans-sialidase, the major Trypanosoma cruzi virulence factor: Three decades of studies,” *Glycobiology*, vol. 25, no. 11. pp. 1142–1149, 2015.

[44] S. Julien *et al.*, “ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumourigenicity,” *Glycobiology*, vol. 16, no. 1, pp. 54–64, 2006.

[45] R. Kannagi, J. Yin, K. Miyazaki, and M. Izawa, “Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants-Hakomori’s concepts revisited,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1780, no. 3. pp. 525–531, 2008.

[46] K. Hatano, Y. Miyamoto, N. Nonomura, and Y. Kaneda, “Expression of gangliosides, GD1a, and sialyl paragloboside is regulated by NF-kappaB-dependent transcriptional control of alpha2,3-sialyltransferase I, II, and VI in human castration-resistant prostate cancer cells.,” *Int. J. cancer*, vol. 129, no. 8, pp. 1838–1847, 2011.

[47] A. Schietinger *et al.*, “A Mutant Chaperone Converts a Wild-Type Protein into a Tumor-Specific Antigen,” *Science (80-.).*, vol. 314, no. 5797, pp. 304–308, 2006.

[48] Y. Kakugawa *et al.*, “Up-regulation of plasma membrane-associated ganglioside sialidase (Neu3) in human colon cancer and its involvement in apoptosis suppression.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 16, pp. 10718–23, 2002.

[49] E. J. Varki A, Cummings RD, “Glycosylation changes in cancer, Essentials of Glycobiology,” *Essentials of Glycobiology*. pp. 1–17, 2009.

[50] J. Rini, J. Esko, and A. Varki, “Glycosyltransferases and Glycan-processing Enzymes,” in *Essentials of Glycoboilogy. 2nd edition.*, 2009, pp. 1–7.

[51] Y. J. Kim and A. Varki, “Perspectives on the significance of altered glycosylation of glycoproteins in cancer,” *Glycoconjugate Journal*, vol. 14, no. 5. pp. 569–576, 1997.

[52] F. Dall’Olio and M. Chiricolo, “Sialyltransferases in cancer,” *Glycoconjugate Journal*, vol. 18, no. 11–12. pp. 841–850, 2001.

[53] M. Lise, C. Belluco, S. P. Perera, R. Patel, P. Thomas, and A. Ganguly, “Clinical Correlations of α2,6-Sialyltransferase Expression in Colorectal Cancer Patients,” *Hybridoma*, vol. 19, no. 4, pp. 281–286, 2000.

[54] M. Amado, F. Carneiro, M. Seixas, H. Clausen, and M. Sobrinho-Simoes, “Dimeric sialyl-Le(x) expression in gastric carcinoma correlates with venous invasion and poor outcome,” *Gastroenterology*, vol. 114, no. 3, pp. 462–470, 1998.

[55] S. Nakamori *et al.*, “Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study.,” *Cancer Res.*, vol. 53, no. 15, pp. 3632–7, 1993.

[56] T. de Vries, R. M. A. Knegtel, E. H. Holmes, and B. A. Macher, “Fucosyltransferases: structure/function studies,” *Glycobiology*, vol. 11, no. 10, p. 119R–128R, 2001.

[57] N. Matsuura *et al.*, “Gene expression of fucosyl- and sialyl-transferases which synthesize sialyl Lewis(x), the carbohydrate ligands for E-selectin, in human breast cancer,” *Int. J. Oncol.*, vol. 12, no. 5, pp. 1157–1164, 1998.

[58] N. Hiraiwa *et al.*, “Transactivation of the fucosyltransferase VII gene by human T-cell leukemia virus type 1 Tax through a variant cAMP-responsive element,” *Blood*, vol. 101, no. 9, pp. 3615–3621, 2003.

[59] Y.-C. Liu *et al.*, “Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 28, pp. 11332–11337, 2011.

[60] Y. Sato *et al.*, “Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein,” *N Engl J Med*, vol. 328, no. 25, pp. 1802–1806, 1993.

[61] M. Demetriou, I. R. Nabi, M. Coppolino, S. Dedhar, and J. W. Dennis, “Reduced contact-inhibition and substratum adhesion in epithelial cells expressing GlcNAc-transferase V,” *J. Cell Biol.*, vol. 130, no. 2, pp. 383–392, 1995.

[62] M. Granovsky, J. Fata, J. Pawling, W. J. Muller, R. Khokha, and J. W. Dennis, “Suppression of tumor growth and metastasis in Mgat5-deficient mice.,” *Nat. Med.*, vol. 6, no. 3, pp. 306–312, 2000.

[63] M. Yoshimura, A. Nishikawa, Y. Ihara, S. Taniguchi, and N. Taniguchi, "Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 19, pp. 8754–8, 1995.

[64] E. P. Bennett, U. Mandel, H. Clausen, T. A. Gerken, T. A. Fritz, and L. A. Tabak, "Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family," *Glycobiology*, vol. 22, no. 6. pp. 736–756, 2012.

[65] D. J. Gill *et al.*, "Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 34, pp. E3152-61, 2013.

[66] M. Dalziel *et al.*, "The Relative Activities of the C2GnT1 and ST3Gal-I Glycosyltransferases Determine O-Glycan Structure and Expression of a Tumor-associated Epitope on MUC1," *J. Biol. Chem.*, vol. 276, no. 14, pp. 11007–11015, 2001.

[67] P. Radhakrishnan *et al.*, "Immature truncated O-glycophenotype of cancer directly induces oncogenic features," *Proc. Natl. Acad. Sci.*, vol. 111, no. 39, pp. E4066–E4075, 2014.

[68] S. Julien *et al.*, "Sialyl-Tn vaccine induces antibody-mediated tumour protection in a relevant murine model," *Br. J. Cancer*, vol. 100, no. 11, pp. 1746–1754, 2009.

[69] C. A. Reis, H. Osorio, L. Silva, C. Gomes, and L. David, "Alterations in glycosylation as biomarkers for cancer detection," *J. Clin. Pathol.*, vol. 63, no. 4, pp. 322–329, 2010.

[70] S. Gilgunn, P. J. Conroy, R. Saldova, P. M. Rudd, and R. J. O'Kennedy, "Aberrant PSA glycosylation—a sweet predictor of prostate cancer," *Nat. Rev. Urol.*, vol. 10, no. 2, pp. 99–107, 2013.

[71] V. R. Zurawski, H. Orjaseter, A. Andersen, and E. Jellum, "Elevated serum CA 125 levels prior to diagnosis of ovarian neoplasia: Relevance for early detection of ovarian cancer," *Int. J. Cancer*, vol. 42, no. 5, pp. 677–680, 1988.

[72] F. Safi, W. Schlosser, G. Kolb, and H. G. Beger, "Diagnostic value of CA 19-9 in patients with pancreatic cancer and nonspecific gastrointestinal symptoms," *J Gastrointest Surg*, vol. 1, no. 2, p. 106–12 OD–1997/03/01, 1997.

[73] F. G. Ebeling *et al.*, "Serum CEA and CA 15-3 as prognostic factors in primary breast cancer.," *Br. J. Cancer*, vol. 86, no. 8, pp. 1217–22, 2002.

[74] K. Noda *et al.*, "Gene expression of alpha1-6 fucosyltransferase in human hepatoma tissues: a possible implication for increased fucosylation of alpha-fetoprotein.," *Hepatology*, vol. 28, no. 4, pp. 944–52, 1998.

[75] R. A. Miller and D. S. Spellman, “Mass spectrometry-based biomarkers in drug development.,” *Advances in experimental medicine and biology*, vol. 806. pp. 341–359, 2014.

[76] B. Adamczyk, T. Tharmalingam, and P. M. Rudd, “Glycans as cancer biomarkers,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1820, no. 9. pp. 1347–1353, 2012.

[77] M. J. Kailemia, D. Park, and C. B. Lebrilla, “Glycans and glycoproteins as specific biomarkers for cancer,” *Analytical and Bioanalytical Chemistry*, vol. 409, no. 2. pp. 395–410, 2017.

[78] Z. Kyselova *et al.*, “Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles,” *Clin. Chem.*, vol. 54, no. 7, pp. 1166–1175, 2008.

[79] W. R. Alley, M. Madera, Y. Mechref, and M. V. Novotny, “Chip-based reversed-phase liquid chromatography-mass spectrometry of permethylated N-linked glycans: A potential methodology for cancer-biomarker discovery,” *Anal. Chem.*, vol. 82, no. 12, pp. 5095–5106, 2010.

[80] J. N. Arnold *et al.*, “Novel glycan biomarkers for the detection of lung cancer,” *J. Proteome Res.*, vol. 10, no. 4, pp. 1755–1764, 2011.

[81] G. Yang *et al.*, “Quantitative glycome analysis of N-glycan patterns in bladder cancer vs normal bladder cells using an integrated strategy,” *J. Proteome Res.*, vol. 14, no. 2, pp. 639–653, 2015.

[82] R. Saldova *et al.*, “Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG,” *Glycobiology*, vol. 17, no. 12, pp. 1344–1356, 2007.

[83] W. R. Alley *et al.*, “N-linked glycan structures and their expressions change in the blood sera of ovarian cancer patients,” *J. Proteome Res.*, vol. 11, no. 4, pp. 2282–2300, 2012.

[84] T. Imre *et al.*, “Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals,” *J. Proteomics*, vol. 71, no. 2, pp. 186–197, 2008.

[85] N. Okuyama *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*, vol. 118, no. 11, pp. 2803–2808, 2006.

[86] Z. Lin *et al.*, “Mass spectrometric assay for analysis of haptoglobin fucosylation in pancreatic cancer,” *J. Proteome Res.*, vol. 10, no. 5, pp. 2602–2611, 2011.

[87] Z. Kyselova *et al.*, “Alterations in the serum glycome due to metastatic prostate cancer,” *J. Proteome Res.*, vol. 6, no. 5, pp. 1822–1832, 2007.

[88] C. R. Borges, D. S. Rehder, and P. Boffetta, “Multiplexed surrogate analysis of glycotransferase activity in whole biospecimens,” *Anal. Chem.*, vol. 85, no. 5, pp. 2927–2936, 2013.

[89] S. S. Pinho and C. A. Reis, “Glycosylation in cancer: Mechanisms and clinical implications,” *Nature Reviews Cancer*, vol. 15, no. 9. pp. 540–555, 2015.

[90] A. Varki, R. Kannagi, and B. P. Toole, “Chapter 44 Glycosylation Changes in Cancer,” in *Cancer*, 2009, pp. 1–11.

[91] G. Lauc, M. Pezer, I. Rudan, and H. Campbell, “Mechanisms of disease: The human N-glycome.,” *Biochim. Biophys. Acta*, vol. 1860, no. 8, pp. 1574–82, 2016.

[92] A. Almeida and D. Kolarich, “The promise of protein glycosylation for personalised medicine,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1860, no. 8, pp. 1583–1595, 2016.

[93] H. J. An *et al.*, “Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer,” *J. Proteome Res.*, vol. 5, no. 7, pp. 1626–1635, 2006.

[94] L. R. Ruhaak, S. Miyamoto, and C. B. Lebrilla, “Developments in the identification of glycan biomarkers for the detection of cancer.,” *Mol. Cell. Proteomics*, vol. 12, no. 4, pp. 846–55, 2013.

[95] L. Hajba, E. Csanky, and A. Guttman, “Liquid phase separation methods for N-glycosylation analysis of glycoproteins of biomedical and biopharmaceutical interest. A critical review,” *Analytica Chimica Acta*, vol. 943. pp. 8–16, 2016.

[96] D. C. Frost and L. Li, “Recent advances in mass spectrometry-based glycoproteomics,” *Adv. Protein Chem. Struct. Biol.*, vol. 95, pp. 71–123, 2014.

[97] Y. Mechref, Y. Hu, A. Garcia, and A. Hussein, “Identifying cancer biomarkers by mass spectrometry-based glycomics,” *Electrophoresis*, vol. 33, no. 12. pp. 1755–1767, 2012.

[98] R. Hennig *et al.*, “Towards personalized diagnostics via longitudinal study of the human plasma N-glycome,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1860, no. 8, pp. 1728–1738, 2016.

[99] X. Verhelst *et al.*, *A glycomics-based test predicts the development of hepatocellular carcinoma in cirrhosis*. 2016.

[100] R. Saldova *et al.*, “Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC,” *J. Proteome Res.*, vol. 13, no. 5, pp. 2314–2327, 2014.

[101] S. Hua *et al.*, “Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers.,” *Analyst*, vol. 136, no. 18, pp. 3663–71, 2011.

[102] S. Hua *et al.*, "Isomer-specific LC/MS and LC/MS/MS profiling of the mouse serum N-glycome revealing a number of novel sialylated N-glycans," *Anal. Chem.*, vol. 85, no. 9, pp. 4636–4643, 2013.

[103] M. Klasić *et al.*, "DNA hypomethylation upregulates expression of the MGAT3 gene in HepG2 cells and leads to changes in N-glycosylation of secreted glycoproteins," *Sci. Rep.*, vol. 6, no. 1, p. 24363, 2016.

[104] D. Kolarich, B. Lepenies, and P. H. Seeberger, "Glycomics, glycoproteomics and the immune system," *Current Opinion in Chemical Biology*, vol. 16, no. 1–2. pp. 214–220, 2012.

[105] M. K. Sethi and S. Fanayan, "Mass spectrometry-based N-glycomics of colorectal cancer," *International Journal of Molecular Sciences*, vol. 16, no. 12. pp. 29278–29304, 2015.

[106] S. Zaare, J. S. Aguilar, Y. Hu, S. Ferdosi, and C. R. Borges, "Glycan Node Analysis: A Bottom-up Approach to Glycomics," *J. Vis. Exp.*, no. 111, 2016.

[107] Y. Hu and C. R. Borges, "A spin column-free approach to sodium hydroxide-based glycan permethylation," *Analyst*, vol. 142, no. 15, pp. 2748–2759, 2017.

[108] T. H. Ho *et al.*, "A multidisciplinary biospecimen bank of renal cell carcinomas compatible with discovery platforms at Mayo Clinic, Scottsdale, Arizona," *PLoS One*, vol. 10, no. 7, 2015.

[109] W. E. Grizzle, E. W. Gunter, K. C. Sexton, and W. C. Bell, "Quality management of biorepositories," *Biopreserv.Biobank.*, vol. 13, no. 1947–5543 (Electronic), pp. 183–194, 2015.

[110] B. A. Katchman *et al.*, "Autoantibody biomarkers for the detection of serous ovarian cancer," *Gynecol. Oncol.*, vol. 146, no. 1, pp. 129–136, 2017.

[111] K. S. Anderson *et al.*, "Autoantibody signature for the serologic detection of ovarian cancer," *J. Proteome Res.*, vol. 14, no. 1, pp. 578–586, 2015.

[112] B. A. Katchman *et al.*, "Proteomic mapping of p53 immunogenicity in pancreatic, ovarian, and breast cancers," *Proteomics - Clin. Appl.*, vol. 10, no. 7, pp. 720–731, 2016.

[113] B. M. Murphy, S. Swarts, B. M. Mueller, P. van der Geer, M. C. Manning, and M. I. Fitchmun, "Protein instability following transport or storage on dry ice," *Nat. Methods*, vol. 10, no. 4, pp. 278–279, 2013.

[114] C. R. Borges *et al.*, "Elevated Plasma Albumin and Apolipoprotein A-I Oxidation under Suboptimal Specimen Storage Conditions," *Mol. Cell. Proteomics*, vol. 13, no. 7, pp. 1890–1899, 2014.

[115] E. S. Baker *et al.*, "Mass spectrometry for translational proteomics: progress and clinical implications," *Genome Med.*, vol. 4, no. 8, p. 63, 2012.

[116] N. L. Anderson and N. G. Anderson, “The Human Plasma Proteome: History, Character, and Diagnostic Prospects,” *Mol. Cell. Proteomics*, vol. 1, no. 11, pp. 845–867, 2002.

[117] J. A. Vasseur, J. A. Goetz, W. R. Alley, and M. V. Novotny, “Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins,” *Glycobiology*, vol. 22, no. 12, pp. 1684–1708, 2012.

[118] Y. Miura and T. Endo, “Glycomics and glycoproteomics focused on aging and age-related diseases - Glycans as a potential biomarker for physiological alterations,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1860, no. 8. pp. 1608–1614, 2016.

[119] K. Miyahara *et al.*, “Serum Glycan Markers for Evaluation of Disease Activity and Prediction of Clinical Course in Patients with Ulcerative Colitis,” *PLoS One*, vol. 8, no. 10, 2013.

[120] D. Vanderschaeghe *et al.*, “GlycoFibroTest is a highly performant liver fibrosis biomarker derived from DNA sequencer-based serum protein glycomics.,” *Mol. Cell. Proteomics*, vol. 8, no. 5, pp. 986–994, 2009.

[121] N. Callewaert, H. Van Vlierberghe, A. Van Hecke, W. Laroy, J. Delanghe, and R. Contreras, “Noninvasive diagnosis of liver cirrhosis using DNA sequencer–based total serum protein glycomics,” *Nat. Med.*, vol. 10, no. 4, pp. 429–434, 2004.

[122] M. Wysoczynski and M. Z. Ratajczak, “Lung cancer secreted microvesicles: Underappreciated modulators of microenvironment in expanding tumors,” *Int. J. Cancer*, vol. 125, no. 7, pp. 1595–1603, 2009.

[123] V. R. Martins, M. S. Dias, and P. Hainaut, “Tumor-cell-derived microvesicles as carriers of molecular information in cancer,” *Curr. Opin. Oncol.*, vol. 25, no. 1, pp. 66–75, 2013.

[124] S. Fontana, L. Saieva, S. Taverna, and R. Alessandro, “Contribution of proteomics to understanding the role of tumor-derived exosomes in cancer progression: State of the art and new perspectives,” *Proteomics*, vol. 13, no. 10–11. pp. 1581–1594, 2013.

[125] G. A. Rabinovich and J. R. Conejo-García, “Shaping the Immune Landscape in Cancer by Galectin-Driven Regulatory Pathways,” *Journal of Molecular Biology*, vol. 428, no. 16. pp. 3266–3281, 2016.

[126] M. K. Sethi, W. S. Hancock, and S. Fanayan, “Identifying N-Glycan Biomarkers in Colorectal Cancer by Mass Spectrometry,” *Acc. Chem. Res.*, vol. 49, no. 10, pp. 2099–2106, 2016.

[127] D. Hanahan and R. A. Weinberg, “Hallmarks of cancer: The next generation,” *Cell*, vol. 144, no. 5. pp. 646–674, 2011.

[128] W. D., “Immunosuppression associated with chronic inflammation in the tumor microenvironment,” *Carcinogenesis*, vol. 36, no. 10. pp. 1085–1093, 2015.

[129] L. R. Ruhaak *et al.*, “Serum glycans as risk markers for non-small cell lung cancer,” *Cancer Prev. Res.*, vol. 9, no. 4, pp. 317–323, 2016.

[130] W. R. Alley and M. V. Novotny, “Glycomic analysis of sialic acid linkages in glycans derived from blood serum glycoproteins,” *J. Proteome Res.*, vol. 9, no. 6, pp. 3062–3072, 2010.

[131] K. R. Reiding, D. Blank, D. M. Kuijper, A. M. Deelder, and M. Wuhrer, “High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification,” *Anal. Chem.*, vol. 86, no. 12, pp. 5784–5793, 2014.

[132] S. Holst *et al.*, “Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues,” *Anal. Chem.*, vol. 88, no. 11, pp. 5904–5913, 2016.

[133] K. Gryska *et al.*, “Inflammatory Cytokines Controlling Branching of N-Heteroglycans of Acute Phase Protein,” in *Glycoimmunology*, A. Alavi and J. S. Axford, Eds. Boston, MA: Springer US, 1995, pp. 239–245.

[134] M. Narisada *et al.*, “Identification of an inducible factor secreted by pancreatic cancer cell lines that stimulates the production of fucosylated haptoglobin in hepatoma cells,” *Biochem. Biophys. Res. Commun.*, vol. 377, no. 3, pp. 792–796, 2008.

[135] J. N. Arnold, R. Saldova, U. M. Abd Hamid, and P. M. Rudd, “Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation,” *Proteomics*, vol. 8, no. 16. pp. 3284–3293, 2008.

[136] R. Saldova, M. R. Wormald, R. a Dwek, and P. M. Rudd, “Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis.,” *Dis. Markers*, vol. 25, no. 4–5, pp. 219–32, 2008.

[137] A. Sarrats *et al.*, “Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis.,” *Proteomics. Clin. Appl.*, vol. 4, no. 4, pp. 432–48, 2010.

[138] A. J. Cagnoni, J. M. Pérez Sáez, G. A. Rabinovich, and K. V. Mariño, “Turning-Off Signaling by Siglecs, Selectins, and Galectins: Chemical Inhibition of Glycan-Dependent Interactions in Cancer,” *Front. Oncol.*, vol. 6, 2016.

[139] S. P. Méndez-Huergo, A. G. Blidner, and G. A. Rabinovich, “Galectins: emerging regulatory checkpoints linking tumor immunity and angiogenesis,” *Current Opinion in Immunology*, vol. 45. pp. 8–15, 2017.

[140] A. a Roberts *et al.*, "Galectin-1-mediated apoptosis in mycosis fungoides: the roles of CD7 and cell surface glycosylation.," *Mod. Pathol.*, vol. 16, no. 6, pp. 543–551, 2003.

[141] S. Tsuboi *et al.*, "A novel strategy for evasion of NK cell immunity by tumours expressing core2 O-glycans," *EMBO J.*, vol. 30, no. 15, pp. 3173–3185, 2011.

[142] S. Tsuboi, S. Hatakeyama, C. Ohyama, and M. Fukuda, "Two opposing roles of O-glycans in tumor metastasis," *Trends in Molecular Medicine*, vol. 18, no. 4. pp. 224–232, 2012.

[143] S. Tsuboi, "Immunosuppressive functions of core2 O-glycans against NK immunity," *Trends Glycosci Glycotechnol*, vol. 25, pp. 117–123, 2013.

[144] S. Tsuboi, "Roles of Glycans in Immune Evasion from NK Immunity," in *Sugar Chains*, Springer, 2015, pp. 177–188.

[145] Y. Suzuki *et al.*, "MUC1 carrying core 2 O-glycans functions as a molecular shield against NK cell attack, promoting bladder tumor metastasis," *Int. J. Oncol.*, vol. 40, no. 6, pp. 1831–1838, 2012.

[146] T. Okamoto *et al.*, "Core2 O-glycan-expressing prostate cancer cells are resistant to NK cell immunity," *Mol. Med. Rep.*, vol. 7, no. 2, pp. 359–364, 2013.

[147] J. Van Rinsum, L. a Smets, H. Van Rooy, and D. H. Van den Eijnden, "Specific inhibition of human natural killer cell-mediated cytotoxicity by sialic acid and sialo-oligosaccharides.," *Int. J. Cancer*, vol. 38, no. 6, pp. 915–922, 1986.

[148] S. H. Itzkowitz and S. H. Itzkowitz, "Mucins Bearing the Cancer-associated Sialosyl-Tn Antigen Mediate Inhibition of Natural Killer Cell Cytotoxicity," *Cancer Res.*, vol. 52, no. 17, pp. 4741–4746, 1992.

[149] J. E. Hudak, S. M. Canham, and C. R. Bertozzi, "Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion," *Nat. Chem. Biol.*, vol. 10, no. 1, pp. 69–75, 2013.

[150] M. A. Recchi, M. Hebbar, L. Hornez, A. Harduin-Lepers, J. P. Peyrat, and P. Delannoy, "Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer.," *Cancer Res.*, vol. 58, no. 18, pp. 4066–4070, 1998.

[151] M. Lise, C. Belluco, S. P. Perera, R. Patel, P. Thomas, and A. Ganguly, "Clinical correlations of alpha2,6-sialyltransferase expression in colorectal cancer patients.," *Hybridoma*, vol. 19, no. 4, pp. 281–286, 2000.

[152] K. Moriwaki *et al.*, "Deficiency of GMDS Leads to Escape from NK Cell-Mediated Tumor Surveillance Through Modulation of TRAIL Signaling," *Gastroenterology*, vol. 137, no. 1, 2009.

[153] T. Iwata *et al.*, “Preoperative serum value of sialyl Lewis X predicts pathological nodal extension and survival in patients with surgically treated small cell lung cancer,” *J. Surg. Oncol.*, vol. 105, no. 8, pp. 818–824, 2012.

[154] S. Mizuguchi *et al.*, “High serum concentrations of Sialyl Lewisx predict multilevel N2 disease in non-small-cell lung cancer,” *Ann Surg Oncol*, vol. 13, no. 7, pp. 1010–1018, 2006.

[155] S. Mizuguchi *et al.*, “Clinical Value of Serum Cytokeratin 19 Fragment and Sialyl-Lewis X in Non-Small Cell Lung Cancer,” *Ann. Thorac. Surg.*, vol. 83, no. 1, pp. 216–221, 2007.

[156] S. Mizuguchi *et al.*, “Serum Sialyl Lewis x and cytokeratin 19 fragment as predictive factors for recurrence in patients with stage I non-small cell lung cancer.,” *Lung Cancer*, vol. 58, no. 3, pp. 369–375, 2007.

[157] D. P. Carbone *et al.*, “VeriStrat® classifier for survival and time to progression in non-small cell lung cancer (NSCLC) patients treated with erlotinib and bevacizumab,” *Lung Cancer*, vol. 69, no. 3, pp. 337–340, 2010.

[158] W. L. Akerley, A. M. Arnaud, B. Reddy, and R. D. Page, “Impact of a multivariate serum-based proteomic test on physician treatment recommendations for advanced non-small-cell lung cancer,” *Curr. Med. Res. Opin.*, vol. 33, no. 6, pp. 1091–1097, Jun. 2017.

[159] F. Grossi *et al.*, “Serum proteomic test in advanced non-squamous non-small cell lung cancer treated in first line with standard chemotherapy,” *Br. J. Cancer*, vol. 116, no. September, pp. 1–8, 2016.

[160] E. Milan *et al.*, “SAA1 is over-expressed in plasma of non small cell lung cancer patients with poor outcome after treatment with epidermal growth factor receptor tyrosine-kinase inhibitors,” *J. Proteomics*, vol. 76, pp. 91–101, 2012.

[161] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2016,” *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, 2016.

[162] L. a G. Ries, J. L. Young, G. E. Keel, M. E. Eisner, Y. D. Lin, and M.-J. D. Horner, “Cancer Survival Among Adults: US SEER PRogram, 1988-2001,” *SEER Surviv. Monogr.*, no. 7, pp. 1988–2001, 2007.

[163] M. Babjuk *et al.*, “EAU guidelines on non-muscle-invasive Urothelial carcinoma of the bladder: Update 2013,” *Eur. Urol.*, vol. 64, no. 4, pp. 639–653, 2013.

[164] S. Goodison, C. J. Rosser, and V. Urquidi, “Bladder cancer detection and monitoring: Assessment of urine- and blood-based marker tests,” *Molecular Diagnosis and Therapy*, vol. 17, no. 2. pp. 71–84, 2013.

[165] E. O. Kehinde, F. Al-Mulla, K. Kapila, and J. T. Anim, “Comparison of the sensitivity and specificity of urine cytology, urinary nuclear matrix protein-22 and multitarget fluorescence in situ hybridization assay in the detection of bladder cancer,” *Scand. J. Urol. Nephrol.*, vol. 45, no. 2, pp. 113–121, 2011.

[166] S. Hafeez and R. Huddart, “Advances in bladder cancer imaging,” *BMC Medicine*, vol. 11, no. 1. 2013.

[167] M. Nagata, S. Muto, and S. Horie, “Molecular Biomarkers in Bladder Cancer: Novel Potential Indicators of Prognosis and Treatment Outcomes,” *Disease Markers*, vol. 2016. 2016.

[168] A.-R. Scheffer, S. Holdenrieder, G. Kristiansen, A. von Ruecker, S. C. Müller, and J. Ellinger, “Circulating microRNAs in serum: novel biomarkers for patients with bladder cancer?,” *World J. Urol.*, vol. 32, no. 2, pp. 353–8, 2014.

[169] A. P. Mitra, D. E. Hansel, and R. J. Cote, “Prognostic value of cell-cycle regulation biomarkers in bladder cancer,” *Semin. Oncol.*, vol. 39, no. 5, pp. 524–533, 2012.

[170] S. J. Yun *et al.*, “Cell-free microRNAs in urine as diagnostic and prognostic biomarkers of bladder cancer,” *Int. J. Oncol.*, vol. 41, no. 5, pp. 1871–1878, 2012.

[171] Y. van Kooyk and G. A. Rabinovich, “Protein-glycan interactions in the control of innate and adaptive immune responses,” *Nature Immunology*, vol. 9, no. 6. pp. 593–601, 2008.

[172] S. Ferdosi *et al.*, “Stage Dependence, Cell-Origin Independence, and Prognostic Capacity of Serum Glycan Fucosylation, β1–4 Branching, β1–6 Branching, and α2–6 Sialylation in Cancer,” *J. Proteome Res.*, p. acs.jproteome.7b00672, 2017.

[173] W. Ansar and S. Ghosh, “C-reactive protein and the biology of disease,” *Immunol. Res.*, vol. 56, no. 1, pp. 131–142, 2013.

[174] S. Shrotriya, D. Walsh, N. Bennani-Baiti, S. Thomas, and C. Lorton, “C-reactive protein is an important biomarker for prognosis tumor recurrence and treatment response in adult solid tumors: A systematic review,” *PLoS One*, vol. 10, no. 12, 2015.

[175] A. Mbeutcha *et al.*, “Prognostic significance of markers of systemic inflammatory response in patients with non–muscle-invasive bladder cancer,” *Urol. Oncol. Semin. Orig. Investig.*, vol. 34, no. 11, p. 483.e17-483.e24, 2016.

[176] L. Zhou, X. Cai, Q. Liu, Z. Y. Jian, H. Li, and K. J. Wang, “Prognostic Role of C-Reactive Protein in Urological Cancers: A Meta-Analysis,” *Sci. Rep.*, vol. 5, 2015.

[177] H. Eggers *et al.*, “Serum C-reactive protein: a prognostic factor in metastatic urothelial cancer of the bladder.,” *Med. Oncol.*, vol. 30, no. 4, p. 705, 2013.

[178] A. J. Hülsmeier, M. Tobler, P. Burda, and T. Hennet, “Glycosylation site occupancy in health, congenital disorder of glycosylation and fatty liver disease,” *Sci. Rep.*, vol. 6, 2016.

[179] J. Hamfjord *et al.*, “Serum N-Glycome Characterization in Patients with Resectable Periampullary Adenocarcinoma,” *J. Proteome Res.*, vol. 14, no. 12, pp. 5144–5156, 2015.

[180] B. C. Jansen *et al.*, “Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS,” *Sci. Rep.*, vol. 6, 2016.

[181] A. Mantovani, P. Allavena, A. Sica, and F. Balkwill, “Cancer-related inflammation,” *Nature*, vol. 454, no. 7203. pp. 436–444, 2008.

APPENDIX A

INCLUSION AND EXCLUSION CRITERIA FOR CERTIFIABLY HEALTHY LIVING KIDNEY DONORS

Mayo Clinic Arizona Guidelines

Inclusion Criteria

All persons that wish to be considered as a potential living donor will be evaluated using the following criteria:

- Age:

  1) Between the ages of 18-70. 2) Those above age 70 will be considered on a case-by-case basis.

- BMI and Blood Glucose:

  Based on "Evaluation of Living Kidney Donor: BMI and Glucose Metabolism Guideline": 1) Impaired Fasting Glucose (FBS > 100 and < 126) or IGT (2 hour BS > 140 and < 200) is a relative contraindication to donation. Patients with family history of diabetes, gestational diabetes and/or metabolic syndrome may be at higher risk than others for developing DM. 2) Prospective donors meeting these criteria require counseling about increased risk of developing DM and its consequences.

- Blood Pressure:

  1) All patients worked up at Mayo Clinic will complete either a 24 hour blood pressure monitor, a blood pressure taken on at least two different occasions, or overnight BP monitoring to be analyzed. 2) Criteria for the diagnosis of Hypertension are including clinic or hypertensive therapy nurse blood pressure > 140/90 mm Hg, ABPM awake period (mean value) > 135/85 mm Hg, ABPM overall (mean value) > 130/80. 3) Donors need to have normal blood pressure; however, hypertension may be acceptable if all the following conditions are met: greater than age 40, Caucasian, GFR meets "Evaluation of

Living Kidney Donor: GFR Protocol", and hypertension controlled with one drug + diuretic.

- Kidney Function:

  GFR Standards need to be followed based on "Evaluation of Living Kidney Donor: GFR Protocol".

- Malignancy:

  Guidelines can be found at "Evaluation of the Living Kidney Donor: Donor Malignancy Guideline".

- Crossmatch/ABO:

  Donors in Arizona and Florida: See "Living Donors Blood Type, Subtype Determination Policy" and "ABO Verification for Living Donors" Policy. Donors in Rochester: See "ABO Blood Group and Other Vital Data Compatibility Verification Guideline".

- Pulmonary Nodules:

  Guideline can be found at "Evaluation of the Living Kidney Donor: Pulmonary Nodules Guideline".

- Stones:

  Guidelines can be found at "Evaluation of the Living Kidney Donor: Donor Nephrolithiasis Guideline".

- Microscopic Hematuria:

  Guidelines can be found at "Evaluation of the Living Kidney Donor: Donor with Microscopic Hematuria Guideline".

- Polycystic Kidney Disease:

  Guidelines can be found at “Evaluation of the Living Donor: Polycystic Kidney Disease Guideline”.

- Psychiatric:

  Guidelines can be found at “Evaluation of the Living Donor: Psychiatric Evaluation policy”.

- Donor Coercion:

  Guidelines can be found at “Evaluation of the Living Donor: Coercion Guideline”.

Exclusion Criteria

- The transplant center may exclude a donor with any condition that, in the hospital’s medical/ethical judgment, causes the donor to be unsuitable for organ donation.

- The transplant center will exclude all donors who meet any of the following exclusion criteria: 1) Is less than 18 years old. 2) Is mentally incapable of making an informed decision. 3) History of HIV. 4) Infectious Disease that can be transmitted through transplantation. 5) Active malignancy, or incompletely treated malignancy. 6) High suspicion of donor coercion. 7) High suspicion of illegal financial exchange between donor and recipient. 8) Evidence of acute symptomatic infection (until resolved). 9) Uncontrolled diagnosable psychiatric conditions requiring treatment before donation, including any evidence of suicidality. 10) Uncontrollable hypertension or history of hypertension with evidence of end organ damage. 11) Diabetes mellitus. 12) Consider on an

individual basis, usually not accepted as donor: non-Caucasian with hypertension, other antihypertensive regimens, family history of hypertensive kidney injury, evidence of end organ damage such as Left Ventricular Hypertrophy (LVH), additional risk factors particularly active smoking.

Additional Information Regarding Living Kidney Donor Selection

Donor Screening

- A potential donor will be screened by the living donor coordinator on the phone or by completing an electronic form.
- In addition, a social work interview is to be conducted for all potential donors if requested by the living donor coordinator.
- At that point, blood type and tissue typing will be obtained and reviewed by the donor team.
- If suitable, potential donor will be scheduled for an evaluation.

Donor Evaluation

Living donor candidate workups are valid for a duration of 18 months after being accepted at selection conference; thereafter, any repeat testing necessary will be determined by the evaluating team at the donor site after which the candidate will again be presented at selection conference.

Workup of the donor will include:

- Labs:

  1) CBC with differential. 2) PT/INR and PTT. 3) Renal Profile (BUN, Creatinine, Electrolytes). 4) Fasting glucose and A1C. 5) Liver function profile. 6) Thyroid Stimulating Hormone. 7) Fasting Lipid Profile. 8) Serum protein electrophoresis for age >60. 9) Oral glucose tolerance test for high risk patients

per the “Evaluation of Living Kidney Donor: BMI and Glucose Metabolism Guideline”. 10) HCG Quantitative blood (female < 55).

- Serologies:

  See “Infectious Disease Protocol”.

- Clearance Studies and 24 hour urine collection:

  1) Spot micro albumin/creatinine ratio. 2) Iothalamate GFR clearance and 24 hour creatinine clearance.

- Urine tests:

  1) Routine urinalysis. 2) Midstream Gram stain and culture. 3) Urine microscopy. 4) Stone risk profile for any donor with history of nephrolithiasis per the “Evaluation of Living Kidney Donor: Donor Nephrolithiasis Guideline”.

- Other tests: Chest x-ray, EKG, Exercise Stress Echo (> 60 years or high cardiac risk) (>50 with hypertension or tobacco use) and/or dobutamine or nuclear stress test if clinically appropriate (may consider for younger patients on case to case basis), Screening for Autosomal Dominant Polycystic Kidney Disease (ADPKD) (for related donors of ADPKD recipient) per the “Evaluation of Living Kidney Donor: Polycystic Disease Guideline”, CT angiography: renal protocol to determine: 1) Whether the kidneys are of equal size. 2) If the kidneys have masses, cysts, or stones. 3) If the kidneys have other anatomical defects. 4) Which kidney is more anatomically suited for transplant.

- Cancer screening per American Cancer Society (ACS) guidelines:

  1) Mammogram (females ≥40 or if h/o breast cancer in pre-menopausal 1st degree relative). 2) Cologuard (≥50 or family history) - first tier for low-risk

patients. 3) Females should have a Pap smear every three years, provided most recent Pap smear was normal. If most recent Pap smear was not normal, follow-up should be according to the recommendations of the GYN service. 4) PSA: Age 50 for men who are at average risk of prostate cancer and are expected to live at least 10 more years, Age 45 for men at high risk of developing prostate cancer. This includes African Americans and men who have a first-degree relative (father, brother, or son) diagnosed with prostate cancer at an early age (younger than age 65), Age 40 for men at even higher risk (those with more than one first-degree relative who had prostate cancer at an early age). 5) Low-Dose CT scan for those at high risk for lung cancer (those who meet all of the following: 55 to 74 years of age, have at least a 30 pack-year smoking history AND are either still smoking OR have quit within the last 15 years. (Note: A pack-year is the number of cigarette packs smoked each day multiplied by the number of years a person has smoked. Someone who smoked a pack of cigarettes per day for 30 years has a 30 pack-year smoking history, as does someone who smoked 2 packs a day for 15 years.)

- Consultations:

  1) Nephrology (different physician from recipient if possible). 2) Living Donor Nurse Coordinator. 3) Social Services (different from recipient social worker if possible). 4) Nutrition. 5) Pharmacy. 6) Transplant Surgeon/Urologist. 7) Consider Transplant Psychiatrist/Clinical Psychology Specialist. 8) Independent Living Donor Advocate.

- Any additional tests, procedures, consults or biopsies needed to determine their candidacy is a part of the donor evaluation until the donor is ruled out as a donor.

Selection Conference

- The final decision to proceed with donation made at the Selection Conference, which may include the following multi-disciplinary team members: transplant surgeons, nephrologists, independent living donor advocate, psychiatrist, social worker, dietitian, pharmacists, financial services and nursing personnel.
- Cases will be presented to the Selection Committee once all appointments completed and all results available.
- Results of testing and assessments will be reviewed by the multi-disciplinary Selection Committee, providing an opportunity for all members to raise concerns and discuss any issues regarding the donor's suitability.
- The living donor’s suitability for donation will be thoroughly documented in the donor’s medical record.
- The decision of the committee will be documented in the donor's medical record and communicated to potential donor by a member of the multidisciplinary team.
- Any exceptions to the selection criteria must be approved by the Living Kidney Donor Selection Committee & the reasons for it thoroughly documented in the patient’s medical record.