Blood Plasma-Based Glycan Nodes as Lung Cancer Markers and the Problem of Biospecimen Integrity in a Multi-Site Clinical Study

by

Yueming Hu

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

Approved April 2019 by the
Graduate Supervisory Committee:

Chad R. Borges, Chair
Alexandra Ros
Xu Wang

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Cancer is a major public health challenge and the second leading cause of death in the United States. Large amount of effort has been made to achieve sensitive and specific detection of cancer, and to predict the course of cancer. Glycans are promising avenues toward the diagnosis and prognosis of cancer, because aberrant glycosylation is a prevalent hallmark of diverse types of cancer. A bottom-up "glycan node analysis" approach was employed as a useful tool, which captures most essential glycan features from blood plasma or serum (P/S) specimens and quantifies them as single analytical signals, to a lung cancer set from the Women Epidemiology Lung Cancer (WELCA) study. In addition, developments were performed to simplify a relatively cumbersome step involved in sample preparation of glycan node analysis. Furthermore, as a biomarker discovery research, one crucial concern of the glycan node analysis is to ensure that the specimen integrity has not been compromised for the employed P/S samples. A simple P/S integrity quality assurance assay was applied to the same sample set from WELCA study, which also afford the opportunity to evaluate the effects of different collection sites on sample integrity in a multisite clinical trial.

Here, 208 samples from lung cancer patients and 207 age-matched controls enrolled in the WELCA study were analyzed by glycan node analysis. Glycan features, quantified as single analytical signals, including 2-linked mannose, α2-6 sialylation, β1-4 branching, β1-6 branching, 4-linked GlcNAc, and outer-arm fucosylation, exhibited abilities to distinguish lung cancer cases from controls and predict survival in patients.

To circumvent the laborious preparation steps for permethylation of glycan node analysis, a spin column-free (SCF) glycan permethylation procedure was

developed, applicable to both intact glycan analysis or glycan node analysis, with improved or comparable permethylation efficiency relative to some widely-used spin column-based procedures.

Biospecimen integrity of the same set of plasma samples from WELCA study was evaluated by a simple intact protein assay (ΔS-Cysteinylated-Albumin), which quantifies cumulative exposure of P/S to thawed conditions (-30 °C). Notable differences were observed between different groups of samples with various initial handling/storage conditions, as well as among the different collection sites.

DEDICATION

To my mom, Xiuling, whose love makes me strong and sees me through everything.

To my husband and best friend, Duo, who shares all my happiness and sadness.

To my son, Max, who can walk by himself now and call me "Mama".

To my family, who support me all the time.

ACKNOWLEDGMENTS

I would like to express my deepest respect and appreciation to my advisor, Dr. Chad R. Borges, for his extraordinary guidance and support during my graduate research and study at Arizona State University. Without his mentorship, it would be impossible for me to accomplish all that I have. And I would give my best to become a decent, warm-hearted and conscientious person like him.

I would also like to thank my committee members, Dr. Alexandra Ros and Dr. Xu Wang, for their helpful guidance over the years. In addition, I would like to thank all the other professors and instructors from School of Molecular Sciences at Arizona State University who helped me along the way.

I thank the Center for Personalized Diagnostics and the Biodesign Institute at Arizona State University for providing the opportunity to work with an excellent group of scientists and advanced facilities.

Additionally, I am thankful to Douglas Rehder, Olgica Trenchevska and Dobrin Nedelkov, for their valuable suggestions and encouragements. I would like to thank all my current and former lab mates and friends, Shadi Ferdosi, Stephanie Thibert, Erandi Kapuruge, Joshua Jeffs, Jesus Aguilar, and Nilojan Jehanathan, for keeping me accompanied and helping me through the PhD life.

Also, I am thankful to all the friends I made here at Arizona State University, for always being there for me and enriching my life.

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

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES xi

LIST OF ABBREVIATIONS xiv

CHAPTER

1 INTRODUCTION 1

1.1 Glycans and Lung Cancer Diagnosis 1

1.1.1 Alteration of Glycan Features in Cancer 1

1.1.2 Glycan markers in the Detection of Lung Cancer 8

1.2 Permethylation of Glycans 10

1.2.1 Historical Development of Permethylation 10

1.2.2 Reaction Mechanism of Permethylation 11

1.3 Plasma/Serum Integrity and *Ex Vivo* Oxidation of Albumin 13

1.3.1 Plasma/Serum Handling and Integrity 13

1.3.2 Harnessing Albumin as a Marker of P/S Integrity 15

Figures 20

2 DIAGNOSTIC AND PROGNOSTIC PERFORMANCE OF BLOOD PLASMA GLYCAN FEATURES IN THE WOMEN EPIDEMIOLOGY LUNG CANCER (WELCA) STUDY 25

2.1 Introduction 25

2.2 Materials and Methods 27

2.2.1 Materials 27

2.2.2 Plasma/Serum Samples 28

2.2.3 Experimental Procedures 30

CHAPTER Page

2.2.4 Data Analysis 32

2.3 Results 33

2.3.1 Altered Glycan Features in Stage I-IV Patients 33

2.3.2 Prominent Early-Stage Alteration 35

2.3.3 Negligible Dependence on Smoking-Status, Age and Histological Type 35

2.3.4 Total Glycosylation and Multivariate Model of Glycan Features 37

2.3.5 Prediction of All-Cause Mortality 37

2.4 Discussion 38

2.5 Conclusion 41

Figures 42

Tables 52

3 A SPIN COLUMN-FREE APPROACH TO SODIUM HYDROXIDE-BASED GLYCAN PERMETHYLATION 62

3.1 Introduction 62

3.2 Materials and Methods 65

3.2.1 Materials and Samples 65

3.2.2 Experimental Procedures 66

3.2.3 Data Analysis 72

3.3 Results and Discussion 73

3.3.1 Optimization of Permethylation Reagent Volumes and Order of Addition for the Spin Column-Free (SCF) Permethylation Procedure 73

CHAPTER Page

3.3.2 Permethylation Efficiency of the SCF Procedure Relative to Spin Column-Based Approaches 76

3.3.3 Analysis of Intact Glycans Following SCF Permethylation 77

3.3.4 Application of the SCF Procedure to Human Blood Plasma 79

3.3.5 Application of the SCF Permethylation Procedure to Glycan Node Analysis of Blood Plasma from Breast Cancer Patients 81

3.4 Conclusion 81

Figures 83

Tables 90

4 BIOSPECIMEN INTEGRITY DISCREPANCY DETECTED BY ΔS-CYS-ALB ASSAY IN A MULTI-COLLECTION SITE LUNG CANCER STUDY 93

4.1 Introduction 93

4.2 Materials and Methods 96

4.2.1 Materials 96

4.2.2 Samples 96

4.2.3 Experimental Procedures 97

4.2.4 Data Processing 99

4.3 Results 100

4.3.1 Effect of Initial Handling Conditions on ΔS-Cys-Alb . 100

4.3.2 ΔS-Cys-Alb at Different Collection Sites 102

CHAPTER Page

4.3.3 Equivalent Time Exposed to Room Temperature 103

4.3.4 Comparison of ΔS-Cys-Alb Between Lung Cancer Patients and Control 103

4.4 Discussion 104

4.5 Conclusion 106

Figures 107

Tables 113

5 CONCLUSION AND FUTURE DIRECTIONS 116

REFERENCES 119

APPENDIX 137

A. COPYRIGHT PERMISSIONS FOR ADAPTIONS OF FIGURES 137

LIST OF TABLES

Table		Page
2.1	Basic Clinical Characteristics and N-values of the WELCA Sample Set	53
2.2	Composition of Three Lung Cancer Sample Sets and Their Sub-cohorts	54
2.3	Statistically Significant Differences between Cohorts within the WELCA Study	55
2.4	Statistically Significant Differences between Cohorts within the WELCA Study (by Endogenous Normalization)	56
2.5	Stage-by-stage ROC Comparison of the Top Performing Glycan Nodes	57
2.6	Comparison of Top Performing Glycan Nodes in Male vs. Female Patients with Early Stage Lung Cancer	58
2.7	Correlation Between Age and the Top Performing Glycan Nodes in the WELCA Cases (All stages) and, Separately, Controls	59
2.8	Comparison of the Top Performing Glycan Nodes in Different Histological Types	60
2.9	Stage-by-stage Comparison of Total Glycosylation with Individual Glycan Feature	61
2.10	Survival Prediction by the Top Performing Glycan Nodes in All Stages, Stage III and IV Combined, Stage III Only and Stage IV Only	62
3.1	Extracted Ions for Disaccharide Gal-4-GlcNAc (± 0.15 Da)	91
3.2	Extracted Ions for Disaccharide GalNAc-3-Gal (± 0.15 Da)	91
3.3	Masses (as $[M+Na]^{+}_{avg}$) of Intact O-glycans from Fetuin	92

Table Page

3.4 Masses (as $[M+Na]^{+}_{avg}$) of Intact N-glycans from Ribonuclease B 92

3.5 Intra- and Interassay Reproducibility of Permethylation Procedures as Applied to Glycan Linkage Analysis of Blood Plasma 93

4.1 Compilation of Studies Describing the Stabilities of Clinical Analytes in Blood Specimens under Various Initial Handling/Storage Conditions .. 114

LIST OF FIGURES

Figure Page

1.1 The Strict Acceptor Substrate Specificity of Glycosyltransferases Is Illustrated by the Human B Blood Group α1-3 Galactosyltransferase ... 21

1.2 Types of N-glycans 22

1.3 Complex O-GalNAc Glycans with Different Cores (Identified by Gray Boxes) 23

1.4 Structures of Representative Glycosphingolipids (GSLs) and Glycoglycerolipids 24

1.5 Substitution Nucleophilic Bimolecular (S_N2) Mechanism of Permethylation Step 25

2.1 Conceptual Overview of the Glycan "Node" Analysis Concept, Which Essentially Consists of Applying Glycan Linkage (Methylation) Analysis to Whole Biofluids 43

2.2 Molecular Overview of the Glycan "Node" Analysis Procedure 44

2.3 Univariate Distribution and ROC Curves for the Six Top Performing Glycan Nodes in the WELCA Study 46

2.4 ROC Curves for Four Top Performing Glycan Nodes in Stage I-IV Within Different Lung Cancer Sets 47

2.5 Connection Between Outer-arm Fucosylation and Smoking Status Within the Control Group 48

2.6 Connection Between Age and a Top Performing Glycan Node, 3,4-linked GlcNAc, in the WELCA Study 49

2.7 Multivariate Logistic Regression Models for Stage I–IV Patients from the WELCA Data Set 50

Figure Page

2.8 Survival Curves for the Six Top Performing Glycan Nodes for All Stages Combined 51

2.9 Survival Curves of the Six Top Performing Glycan Nodes for Stage III and IV Combined 52

3.1 Conceptual Overview of Glycan Linkage Analysis from Whole Biospecimens 84

3.2 Molecular Overview of the Glycan Linkage Analysis Procedure 85

3.3 Effect of Original-aqueous-sample Volume on Permethylation Efficiency 86

3.4 Evidence That the New SCF Procedure Has Comparable Permethylation Efficiency for Terminal/non-terminal Hexose and Terminal/non-terminal HexNAc Relative to Other Prevalently Used Spin Column Procedures . 87

3.5 Permethylation of Intact Glycans by the SCF and Extended Spin Column Procedure (SC 50 min) 88

3.6 Ratio of Intact, Permethylated Glycan Yields Produced by the SCF Procedure Relative to a Commonly Employed SC Procedure (SC 50 min) 89

3.7 Application of the SCF Procedure to Glycan Node Analysis of Clinical Samples 90

4.1 Charge Deconvoluted ESI-Mass Spectra of Albumin That Illustrate ΔS-Cys-Alb 108

4.2 Effect of Storage Temperature and Pre-storage Delay on ΔS-Cys-Alb .. 109

4.3 Correlation of Pre- and Post-centrifugation Delay, Total Pre-storage Delay with ΔS-Cys-Alb 110

Figure Page

4.4 Univariate Distribution of ΔS-Cys-Alb at the Different WELCA Study Collection Sites, and from Different Collection Batches at Site 18 111

4.5 Estimate of Plasma Sample Exposure Times to the Equivalent of 23 °C Based on the Rate Law for Formation of S-Cys-Alb 112

4.6 Univariate Distribution of S-Cys-Alb and ΔS-Cys-Alb for Lung Cancer Patients and Controls 113

LIST OF ABBREVIATION

ACN	Acetonitrile
ACTH	Adrenocorticotropic hormone
AGP	α_1-acid glycoprotein
APTT	Activated partial thromboplastin time
Asn	Asparagine
AUC	Area under curve
CL	Cell line
CL	Confidence level
CT	Computed tomography
CTX	C-telopeptides of type I collagen
Cys-SOH	Cysteine sulfenic acid
DHB	2,5-Dihydroxybenzoic acid
Dimsyl	Methylsulfinyl carbanion
DMSO	Dimethyl sulfoxide
DOPAC	Dihydroxyphenylacetic acid
DOPEG	3,4-dihydroxyphenylglycol
EGF	Epidermal growth factor
ER	Endoplasmic reticulum

ESI	Electrospray ionization
Fc	Fragment crystallizable
FGF-2	Fibroblast growth factor-2
FSH	Follicle-stimulating hormone
FTCs	Freeze-thaw cycles
Fuc	Fucose
FUT8	α1,6-fucosylatransferase
Gal	Galactose
GalNAc	N-acetylgalactosamine
GC-MS	Gas chromatography-mass spectrometry
GDP	Guanosine diphosphate
Glc	Glucose
GlcNAc	N-acetylglucosamine
GSLs	Glycosphingolipids
GTs	Glycosyltransferases
5-HETE	5-hydroxyeicosatetraenoic acid
HexNAc	N-acetylhexosamine
HS	Heparan sulfate
HSA	Human serum albumin

Ig	Immunoglobulin
LDCT	Low dose chest computed tomography
MALDI	Matrix-assisted laser desorption/ionization
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
Man	Mannose
MeI	Methyl iodide
MeOH	Methanol
MHC	Major histocompatibility complex
MMP	Matrix metalloproteinases
MS	Mass spectrometry
Na dimsyl	Sodium methylsulfinyl carbanion
NaH	Sodium hydride
NaOH	Sodium hydroxide
NAs	Normalized abundances
NCI	National Cancer Institute
NeuNAc	N-acetylneuraminic acid, aka sialic acid
NLST	National Lung Screening Trial
NSCLC	Non-small cell carcinomas
NSCLC	Non-small cell lung cancer

-OH	Hydroxyl group
P/S	Plasma or serum
PAVs	Pre-analytical variables
PMAAs	Partially methylated alditol acetates
PNGase F	Peptide-N-Glycosidase F
PSA	Prostate specific antigen
PTH	Parathyroid hormone
PTM	Post-translational modification
QC	Quality control
QC/QA	Quality control/quality assurance
ROC	Receiver operating characteristic
ROS	Reactive oxygen species
S-Cys-Alb	S-cysteinylated albumin
S/N	Signal-to-noise
SC	Spin column
SCF	Spin column-free
SCLC	Small cell lung carcinomas
SEM	Standard error of the mean
Ser	Serine

SLe^x	Sialyl-Lewis^x
S_N2	Substitution nucleophilic bimolecular
SOPs	Standard operating procedures
t-Gal	Terminal galactose
TFA	Trifluoroacetic acid
TGF	Transforming growth factor
Thr	Threonine
TIC	Total ion current chromatogram
TLR4	Toll-like receptor 4
UDP	Uridine diphosphate
VEGF	Vascular endothelial growth factor
WELCA	Women Epidemiology Lung Cancer
XIC	Extracted-ion chromatogram
Xyl	Xylose

CHAPTER 1

INTRODUCTION

1.1 Glycans and Lung Cancer Diagnosis

Glycans are complex biological sugar polymers that are usually attached to proteins and lipids and are often displayed on cell surfaces. Like nucleic acids, proteins and lipids, glycans are primary constituents of cells and play significant roles in many vital biological processes. The biological functions of glycans span a wide range: 1) structural and modulatory functions, such as facilitating protein folding; 2) specific recognition by intrinsic and extrinsic glycan-binding proteins, such as those involved in signal transduction; 3) cellular adhesion; and 4) molecular mimicry of host glycans. But among the four major classes of biomolecules, the analysis of glycans is the most challenging, mainly due to the abundance and enormous structural diversity of glycans as well as the fact that they are not constructed in a template-driven manner. Nevertheless, in the past few decades, impressive progress has been achieved in the development of technologies, especially in mass spectrometry, that benefit glycan analysis[1]. The advances in ionization technologies[2, 3], evolved methods of ion activation[4, 5], improved chromatographic separations[6-8], and introduction of ion mobility mass spectrometry[9] all expedite the characterization of glycans. Glycomics is the comprehensive study of glycans, in the forms of oligosaccharide polymers, N- and O-linked glycoproteins, glycolipids, and proteoglycans. Glycosylation describes the process of attaching glycans to proteins or other biological molecules in endoplasmic reticulum (ER) and golgi apparatus organelles. And the altered glycosylation has long been discovered as a universal feature of cancer cells[10].

1.1.1 Alteration of Glycan Features in Cancer

1.1.1.1 Heterogeneity and Consistency of Glycan Construction

Glycosylation is the most common and well-known post-translational modification (PTM) reaction. It is catalyzed by enzymes called glycosyltransferases (GTs). Glycosyltransferases assemble monosaccharide moieties to form linear or branched glycan chains, by transferring and attaching the monosaccharide moiety of a donor substrate to the acceptor substrate, using an S_N2 (substitution nucleophilic bimolecular) reaction mechanism[11]. The most common glycosyl donors in eukaryotic cells are nucleotide sugars, such as Uridine diphosphate (UDP)-glucose and Guanosine diphosphate (GDP)-mannose. And acceptors are commonly oligosaccharides, proteins, lipids and sometimes DNA.

Unlike the synthesis of DNA, RNA and proteins, there is no biologically embedded template for glycan construction. Despite the fact that most human glycans are comprised of only nine major monosaccharides, the large family of GTs together with their non-template-driven manner lead to the natural complexity and structural heterogeneity of glycans. However, among the seemingly chaotic process, the strict acceptor, donor, and linkage specificity[11] of GTs contribute to moderate degree of consistency of glycan construction. In general, each glycosyltransferase catalyzes the formation of a single glycosidic linkage[12, 13]. An excellent example of the specificity can be provided by human B blood group α1–3 galactosyltransferase, as illustrated in **Fig. 1.1**. The B transferase adds galactose (from its specific donor UDP-galactose) in unique α1–3 linkage to H antigen, the specific acceptor of this enzyme. The strict acceptor substrate specificity of B transferase lies in the fact that it only acts on the α1–2-linked fucose modified H antigen, but not on the unmodified type-2 precursor or other modified form of the precursor. Exceptions to the strict donor, acceptor, and

linkage specificity exist in instances that more than one GT share the same acceptor and make the same linkage, such as human fucosyltransferases III–VII[14]. Other examples include GTs that catalyzes multiple reactions (β1-4 galactosyltransferase[15-17]) and those with two separate active sites (such as those that synthesize the backbones of glycosaminoglycans). Nevertheless, most GTs exhibit the specificity strictly and thus define and limit the number and type glycan structures present in a given organism[11].

1.1.1.2 N-glycans, O-glycans and Glycolipids

The three major forms of glycans prevalent in most eukaryotes are N-glycans, O-glycans and glycolipids. (Abbreviations of commonly seen monosaccharide moieties are provided above.)

N-glycans are covalently linked to the nitrogen in the side chain of asparagine (Asn) by an N-glycosidic bond. A sequon beginning with asparagine followed by any amino acid except proline and ending with serine or threonine (Asn-X-Ser/Thr) is a requisite for a protein to receive N-glycans. A common core structure, Man α1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn-X-Ser/Thr, is shared by all N-glycans. N-glycans are classified into three types based on the composition of their "antennae" attached to the core sequence (**Fig. 1.2**): 1) oligomannose, with the antennae composed of only mannose residues; 2) complex, consisting of antennae that add a GlcNAc residue (usually followed by Galactose and Sialic acid attachment) to the Manα1-3 and Manα1-6 arms; 3) hybrid, with mannose residues linked to the Manα1-6 arm and one or two antennae initiated with GlcNAc residue linked to the Manα1-3 arm.

Our understandings of the N-glycans biological functions have been enormously advanced in the past few decades. The research reported by Ari Helenius and colleagues elucidated the crucial function of the N-glycans in facilitating the proper folding of proteins[18, 19]. In brief, calnexin and calreticulin in ER, which serve as chaperones for the newly synthesized glycoproteins during their folding so that they do not aggregate, or leave the ER prematurely, bind specifically to the three glucose residues of the N-glycans attached to a nascent polypeptide[20, 21]. The three glucose residues will be trimmed if the glycoprotein has properly folded, otherwise reattachment of the three glucose residues onto the N-glycan structure and further folding processes will subsequently occur. Additionally, N-glycans play key roles in signaling. As revealed by Wang et al., deficiency of the core fucosylation down-regulated several receptor-mediated signaling pathways, including transforming growth factor (TGF)- β1 receptor, epidermal growth factor (EGF) receptor, and integrins, which are responsible for cell growth, differentiation, and emphysema[22]. Moreover, N-glycans present in the fragment crystallizable (Fc) region of immunoglobulin (Ig) facilitate the dictation of the antibody's effector function[23].

O-glycans are covalently linked to hydroxyl group (-OH) of serine (Ser) or threonine (Thr) by an O-glycosidic bond. Unlike N-glycans, no consensus sequence of the receiving peptide chain for O-glycans is discovered yet. There are several types of O-glycans, among which the α-linked O-GalNAc glycans from mucin are the most ubiquitous and crucial. Four common core structures, which can be subsequently extended to form mature linear or branched O-GalNAc glycans, are illustrated in **Fig. 1.3**.

Many essential biological functions of O-glycans have been discovered. The mucin O-GalNAc glycans are essential in promoting mutualism with the commensal

microbiota in the intestinal tract[24]. The core 1-derived O-GalNAc glycans[25] are pivotal in sustaining the protective homeostatic mucus barrier between the resident microbiota and the underlying immune cells in the colon. Meanwhile, lack of the core 3-derived O-GalNAc[26] can lead to physiological defects, including increases in intestinal permeability and higher levels of colonic bacteria within the colonic mucosa. Another important O-glycan, Sialyl-Lewis^x (SLe^x) plays vital role in cell-cell interaction[27] and serves as an essential blood group antigen. Furthermore, SLe^x is critical to immune and inflammatory response[28].

Glycolipids are lipids covalently attached by glycans with a glycosidic bond. Glycosphingolipids (GSLs) and glycoglycerolipids are the two major types of glycolipids, that are glycans built on a ceramide lipid moiety that consists of a long-chain amino alcohol (sphingosine) in amide linkage to a fatty acid, and a diacyl or acylalkylglycerol lipid moiety, respectively[29]. Besides the β-linked galactosylceramide (GalCer) and its analog sulfatide that are prevalent especially in brain, most complex vertebrate glycosphingolipids share a fundamental core structure lactosylceramide (Galβ1-4GlcβCer). Structures of representative glycolipids for the two types and one example of a complex sialylated GSL are provided in **Fig. 1.4**.

Complex glycosphingolipids function both in cell-cell recognition and in the regulation of signal transduction[30], as their *trans* recognition roles. When it comes to the *cis* recognition role, they also interact laterally with proteins in the same membrane to modulate their activities[31]. The simple glycolipid GalCer functions as obligate precursor to the ceramide required to build the outermost protective layer (stratum corneum) of the skin[32]. And its analog sulfatide have essential roles in myelin-axon interactions, the absence of which leads to serious neurological deficits[33].

1.1.1.3 Changes of Glycans in Cancer and the Associated Functions

Altered glycosylation is a universal hallmark of various cancer types. There are a few different forms of glycan alteration in malignant cells: increased expression of and decreased expression of certain glycans, overexpression of truncated and incomplete glycans, and emergence novel glycans[10]. The changes of several common glycan forms are listed and discussed below.

Many studies have revealed the changes of N-glycans in a variety of cancer types, as well as their significant functions in facilitating the metastasis and progression of cancer. Increased β1-6 branching N-glycans, resulted from excessive expression of N-acetylglucosaminyltransferase V (GlcNAc-TV, or MGAT5) and the upstream enhanced transcription of the MGAT5 gene, were suggested to expedite the metastasis process of tumor cells[10, 34-36]. As reported by Seberger et al.[36] in 1999, the transfected mouse mammary cancer cell lines with increased β1-6 branching on the cell surface formed elevated levels of lung tumors with 4- to 40-fold in metastatic potential in mice relative to control transfected cell lines. Another persuasive evidence comes from the research performed by Granovsky et al[35]. that revealed suppression of tumor growth and metastasis in MGAT5-deficient mice. Besides, increased β1-4 branching tetra-antennary N-glycans, arisen from the overexpression of MGAT4 gene, also contribute to tumor progression[10, 37]. The increased N-glycan branching, together with enhanced out-arm fucosylation[38], give rise to tumor metastasis and progression, probably by the generation of poly-Nacetyllactosamines (LacNAc) and SLe^x (ligands for galectin and selectin, respectively) and subsequently extended growth-factor signaling[10, 37].

The altered O-glycans in cancer also exhibit crucial functions. Mucins, which carry plenty of O-GalNAc glycans, were disclosed to promote the metastasis of tumor

cells and inhibit the activity of cytolytic cells such as natural killer cells[39]. Due to the rod-like structure and negative charge, mucins are capable to repel intercellular interactions, for instance between blood-borne tumor cells and host cytolytic cells[40, 41]. Also, the structural property enables mucins to sterically inhibit the activities of other adhesion molecules as cadherins and integrins, and lead to the initiation of tumor metastasis[42, 43]. Moreover, the antigen peptides present on the surface of major histocompatibility complex (MHC) may be blocked by mucins[10]. The carcinoma mucins feature by their incomplete glycan forms, primarily Tn (GalNAcα1-O-Ser/Thr), T (Galβ1-3Tn) and sialyl Tn (NeuNAcα1-6Tn, sTn) antigens[44], that are correlated with silencing of the COSMC gene (C1GALT1C1) and overexpression of the sialyltransferase ST6GALNAC1.

Some glycolipids are largely overexpressed in specific cancers. For instance, GD2 and GD3, disialogangliosides highly enriched in on tumors of neuroectodermal origin, are involved in the attachment of human melanoma and neuroblastoma cells to extracellular matrix proteins, as suggested by Cheresh et al[45]. Yet the modulatory mechanism of gangliosides remains obscure. Fishman and colleagues[46] had reported gangliosides as a bimodal regulator of cell growth, which may either promote or inhibit cell growth depending on the state of growth of the target cell.

Altered expression of sialic acid is another representative and early-observed hallmark of cancers. Increased α2-6 sialylation on the outer LacNAc units, resulted from the upregulation of ST6GAL1gene, was indicated to promote tumor progression[10, 47], by enhancing integrin[48, 49] activity and cell motility[49, 50], and suppressing galectin-3-induced apoptosis[51] and Fas death receptor activity[52]. However, the modulatory role of α2-6 sialylation in cancer is controversial, with some

studies discovered that it also inhibits tumor progression[53] by hindering the binding between epidermal growth factor (EGF) and EGF receptor.

Heparan sulfate (HS) is a unique sulfated glycosaminoglycan (GAG) that has been implicated in tumor pathogenesis[10, 54], primarily by promoting invasion[55, 56], metastasis[56] and angiogenesis[57]. HS covalently linked proteoglycans are found to facilitate cell-cell and cell-matrix interactions[58] and bind numerous growth factors, such as fibroblast growth factor-2 (FGF-2)[59] and vascular endothelial growth factor (VEGF)[60], and a variety of other chemokines[61] and cytokines[62].

1.1.2 Glycan Markers in the Detection of Lung Cancer

Up to now, the greatest number of cancer death in United States comes from lung cancer, with an estimate of 142,670 deaths in 2019[63]. Lung cancer is categorized into two main histological types, small cell lung carcinomas (SCLC) and non-small cell carcinomas (NSCLC). The second histology type covers more than 80% of lung cancer and consists of three subgroups: adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Currently, the detection and diagnosis of lung cancer largely rely on chest X-ray films and computed tomography (CT) scans. The lack of early stage symptoms and diagnostic technologies with acceptable sensitivity and specificity give rise to less than 5% five-year survival rate of lung cancer[63]. Lung cancer is known to closely associated with various glycosylation changes[64], which can potentially serve as glycan-based markers to aid in lung cancer diagnosis, especially in early stage.

It has been reported by Zhang et al. in 1997 that lung cancer cells upregulated the expression of sialylated and fucosylated glycan epitopes on the cell surface[65, 66], including SLe^x, SLe^a, sTn, Le^y, gangliosides GM2, fucosyl GM1, Globo H, polysialic acid, etc. Further studies focusing on uncovering the link between glycosylation and

protein functions were carried out by Liu et al.[67] using the more recent technologies-alkyne-sugar probe and Cu [I]-catalyzed alkyne-azide click chemistry[68]. By comparing the sialylation and fucosylation levels of two lung cancer cell lines derived from same parental cell line but with distinct invasiveness (CL1-5, high invasive and CL1-10, low invasive), Liu and colleagues identified greatly higher expression level (22.5 folds) of FUT8 gene in the move invasive lung cancer cell line (CL1-5). The overexpressed α1,6-fucosylatransferase (FUT8) and corresponding core fucosylation glycan feature was found to be essential to EGFR dimerization and phosphorylation, and the consequential binding affinity to EGF[67].

Lung cancer and smoking-induced changes in N-glycosylation of blood serum proteins were evaluated by Vasseur et al, with the discovery of elevated level of fucosylated tetra-antennary structures with varying degrees of sialylation in lung cancer patients. Further investigation revealed that outer-arm fucosylation, particularly for the tri- and tetra-antennary structures, was elevated in both lung cancer patients and controls with previous smoking history. Besides, increased α2,6-linked sialic acids isoform and decreased α2,3-linked sialic acid isoform were also observed in former smokers[69]. Recently, Arnold et al. reported several novel glycan biomarkers for lung cancer detection, including increases of SLe^x, monoantennary glycans, highly sialylated glycans and decreases of core-fucosylated biantennary glycans, with some being detectable as early as in Stage I[70]. In 2018, Ferdosi et al.[71, 72] evaluated the detection and prediction performance of plasma and serum glycomics in lung, prostate, serous ovarian, pancreatic and bladder cancer, indicating stage dependent, cell-origin independent diagnostic and prognostic capacity of serum glycan fucosylation, α2-6 sialylation, β1-4 branching, β1-6 branching in lung and bladder cancer.

1.2 Permethylation of Glycans

The unprecedented progress in the development of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) MS has substantially expedited structural analysis of glycans in the last two decades. Although most glycans can be directly analyzed by MALDI-MS in their native form, several pronounced advantages can be conferred to the structural analysis of glycans by permethylation[73]: 1) the methyl groups added to all the hydroxyl and N-acetyl groups within oligosaccharide polymer during permethylation increase the hydrophobicity of glycan, which facilitates the separation by LC-MS; 2) sialic acid, as the most labile monosaccharide moiety, can be stabilized by methylation, with the carboxyl functional group methyl esterified; 3) the ionization efficiency and quantitative reproducibility of glycans on MS (especially MALDI- and ESI-MS) can be considerably enhanced; 4) determination of branching and glycosidic linkage positions can be easily achieved[74, 75]; 5) more predictable spectral patterns can be generated for tandem mass spectral analysis (MS/MS)[76].

1.2.1 Historical Development of Permethylation

The development history of permethylation procedure spanned over a century[77]. The first glycan methylation procedure was developed by Prudie and Irvine in 1903[78], in which carbohydrates were dissolved in methanol, reacted with methyl iodide (MeI) in the presence of silver oxide. About a decade later, Haworth et al.[79] introduced methyl sulfate as the methylation reagent and sodium hydroxide solution to react with carbohydrates dissolved in trace amount of water. To improve the undesirable methylation yields from these two procedures, in 1960s[80-83], the methylsulfinyl carbanion (dimsyl) and dimethyl sulfoxide (DMSO) were introduced as catalyst and

solvent respectively, with methyl iodide working as methylation reagent. Nevertheless, even with obviously enhanced methylation yield and achievement of permethylation of carbohydrates in one step, the preparation of dimsyl reagent, which is extremely sensitive to air, moisture and carbon dioxide, remained laborious and hazardous. In 1984, Ciucanu and Kerek utilized powdered sodium hydride (NaH) to substitute Na dimsyl, with methyl iodide and DMSO remained in the procedure[84]. Yet, this procedure that was widely adopted by many glycan studying laboratories required quite large quantity of carbohydrates. Subsequently, in 2005, Mechref, Novotny and colleagues[85] developed a solid-phase spin column-based procedure, where carbohydrates dissolved in DMSO were proposed to react with methyl iodide on the surface of NaOH beads. This procedure was adopted and optimized by Borges group[74, 75], and employed to various types of cancer sets.

1.2.2 Reaction Mechanism of Permethylation

The mechanism of NaOH beads-based permethylation procedure optimized by Borges group is described below. The solvent used to dissolve glycan sample solution as well as plasma or serum samples is DMSO with trace amount of water. The NaOH beads are responsible for generating alkaline condition. As depicted in **Fig. 1.5**, the permethylation reaction is carried out in an S_N2 mechanism. Firstly, all free hydroxyl groups within the polysaccharide structure are deprotonated by NaOH, to generate alkoxide ions (RO^-). Next, the alkoxide ions work as powerful nucleophiles and attack the electrophilic carbon in methyl iodide, with the leaving group iodine ion leaving and the methyl group covalently linked to the alkoxyl group. The yield of this permethylation step with 24% w/w of NaOH/DMSO condition is about 99%.[86, 87] Thus, even though the undermethylated byproducts of permethylation step may interfere with

the subsequent analysis and lead to biased result, the undermethylation of polysaccharides can often be minimized, with negligible effects on the quantitative determination of glycosidic linkage positions.

1.3 Plasma/Serum Integrity and *Ex Vivo* Oxidation of Albumin

In the past few decades, the remarkable progresses in the development of automation instrument and computer sciences, accompanied by the implementation of high-quality analytical standards, have substantially improved the quality of lab testing results and diminished analytical errors. Hence, more attentions have been focused on errors incurred in the pre-analytical phase, primarily due to the fact that it is currently difficult to monitor all the preanalytical variables (PAVs), especially when they are not under direct laboratory supervision[88, 89]. Compromised integrity of biospecimens should pertain to the loss of an ability to accurately quantify one or more analytes—i.e, a sample's "integrity" depends on what analytes are going to be analyzed and whether or not they've remained stable. PAVs that largely impact the integrity of biospecimens pervade in the following four steps: collection, processing, transport and storage. Notably, the research presented in this dissertation is primarily concerned with handling and storage conditions that result in the exposure of specimens to thawed conditions—e.g., for blood plasma and serum, temperatures > -30 °C[90, 91]—because this is often the most difficult PAV to control and track over the lifetime of a specimen collected and archived for research purposes.

1.3.1 Plasma/Serum Handling and Integrity

Among a variety of biospecimens that are commonly used in clinical chemistry laboratories, plasma/serum (P/S) samples stand out with the prominent advantage that most analytes of interest are available in blood biospecimen. Even with some disadvantages due to mildly painful collection, blood samples account for more than half of the biospecimens included in the National Cancer Institute (NCI)-sponsored projects[92].

PAVs involved in the *collection* step of P/S samples are brands of collection tube, additives added in collection tube (EDTA, heparin, sodium fluoride/potassium oxalate), types of specimens (whole blood, plasma, or serum), mixing of blood, hemolysis, and so on. In 2001, Evans et al.[93] tested the effect of several prevalent additives used in blood collection on stability of plasma and serum hormones. They recommended the utilization of EDTA as the anticoagulant added in blood collection tube when hormones are the analytes of interest, because both heparin and fluoride significantly affected concentrations of some hormones relative to EDTA.

Within the *processing* step of P/S samples, considerable alterations of sample integrity could come from many important PAVs, such as speed, gravity, temperature and duration of centrifugation. Reisch et al.[94] suggested that preanalytical stability of adrenocorticotropic hormone (ACTH) relied more on time to centrifugation than temperature, with significantly decreased ACTH concentrations in uncentrifuged samples compared to centrifuged samples stored at same temperature (4 ℃). According to Tanner and colleagues[95], delay in centrifugation at various temperatures up to 24 hours significantly affected some analytes such as potassium, urea, glucose, iron, and ferritin. And generally, temperature-controlled centrifuges are recommended[96].

Essential PAVs for *transport* and long-term *storage* of P/S samples include exposure to thawed conditions (> -30 ℃ [90, 91]) probably due to loss of dry ice, storage duration and temperature, and numbers of freeze-thaw cycles (FTCs). Holland et al.[97] have reported that repeated freeze-thaw cycles can be detrimental for some analytes in blood samples, such as RNA. Therefore, aliquots should be created upfront rather than delayed, to avoid repeated FTCs. The storage temperature also matters for blood sample integrity. A variety of plasma/serum (P/S) analytes were found to be unstable when

stored in -20 ℃ for longer than 3 months, such as VEGF-receptor[98], thromboplastin and prothrombin[99], etc. Even stored at -80 ℃, some analytes were revealed to be unstable after a certain storage duration. For example, matrix metalloproteinases (MMP)-9[100] was found to loss stability within one-month storage in -80 ℃, while apolipoprotein E[101] was reported to be stable for only 3 months in -80 ℃.

To circumvent the issue of fluctuated P/S sample integrity due to so many PAVs, substantial efforts have been made to accelerate the development of standard operating procedures (SOPs), that guide the collection, processing, transport and storage conditions for blood specimens. Nevertheless, a rigid SOP—while appealing to administrators—may be difficult and impractical to follow for some collection sites, because of possible logistical and financial restrictions. Furthermore, some PAVs and conditions are not easy to keep track of and document in detail, such as delay in shipments and failure of a freezer. Hence, an accurate endogenous marker to assure the integrity of P/S samples, especially for those that are pre-existing or lack detailed paper trials, is essential and beneficial. Recently, the Borges group[102] has developed a "ΔS-cysteinylated-albumin" (ΔS-Cys-Alb) assay, which measures the ex-vivo oxidizability of human serum albumin (HSA) and applies it as a QC/QA marker of P/S integrity.

1.3.2 Harnessing Albumin as a Marker of P/S Integrity

1.3.2.1 Structure and Biological Functions of Human Serum Albumin

Human serum albumin (HSA) is the most abundant circulating protein in human P/S, which accounts for over 50% of total plasma proteins. The composition and structure of HSA has been well understood by scientists[103, 104]. The native HSA is a

non-glycosylated protein composed by 585 amino acids and 3 domains, with a molecular weight of 66438 Dalton (Da). There are a total 35 cysteine residues in HSA sequence, forming 17 disulfide bonds and nine (practically eight and half) double loops. The double loops are further grouped into three homologous domains, each of which consisting of 2 longer loops with a shorter loop in between. In each domain, the first two loops are grouped as subdomains A (IA: loop 1-2; IIA: loop 4-5; IIIA: loop 7-8) and the third loop is referred as subdomain B (IB: loop 3; IIB: loop 6; IIIB: loop 9). The hydrophobic cavities in subdomains IIA and IIIA are the primary ligand binding regions of HSA. The cysteine residue at position 34 (Cys34) of HSA is primarily free and can react with other circulating molecules.

HSA has many essential biological functions, such as maintaining osmotic pressure, binding to endogenous and exogenous compounds, balancing the redox state, providing inflammatory and/or immunological responses, and regulating the pharmacokinetic and pharmacodynamics of many drugs[104, 105]. As the predominate plasma protein with such a large molecular weight, HSA acts as the main modulator of plasma oncotic pressure[106]. The negative net charge of HSA at physiological pH also contribute to its oncotic capacity by Gibbs-Donnan effect-i.e., by attracting positively charged counter-ions to the intravascular compartment[104]. In addition, the binding sites located in subdomains IIA and IIIA enable the binding of HSA to fatty acids, hormones, ions and drugs, which consequently increases the solubility of the bound compounds and favors their distributions to the action sites[107]. Furthermore, the antioxidant activity of HSA comes from the free cysteine residue Cys34 that is capable to bind to reactive oxygen species (ROS)[108, 109], and the six methionine residues which can be oxidized to sulfoxides[110-112]. The N-terminus of HSA is also adjuvant to its antioxidant capacity, by

chelating redox-active transition metal ions and preventing them from catalyzing the formation of hydroxyl radicals[111, 113]. Moreover, HSA binds to surface components of gram-negative and gram-positive bacteria that activate the innate immune system through Toll-like receptor 4 (TLR4) and induce inflammation[114-116]. Besides, binding to HSA controls the free, active concentrations of many drugs and may affect considerably their overall pharmacodynamic and pharmacokinetic profile[117-119].

1.3.2.2 *Ex Vivo* Oxidizability of Albumin as a Marker of P/S Integrity

Some evidence exists that albumin is susceptible to *ex vivo* oxidation[120]. Recently, Borges group has performed systematic studies of the *ex vivo* oxidation of albumin[102, 121, 122]. It has been well known that free thiol-containing molecules, such as albumin, will spontaneously form intra- or intermolecular disulfide bonds in air-exposed aqueous solutions, in the absence of special precautions. As shown by numerous groups over the past few decades[123-125], this spontaneous process only requires atmospheric oxygen and trace metals and, as shown by the Borges group, occurs, in part, through a cysteine sulfenic acid (Cys-SOH) intermediate[121]. The S-cysteinylation of albumin through disulfide exchange with cystine (the dimerized form of cysteine) present in P/S, also occurs and appears to be the dominant mechanism governing the ex vivo oxidative S-cysteinylation of albumin. The reactions governing the *ex vivo* formation of S-cysteinylated albumin (S-Cys-Alb) are as follows:

$$AlbSH + Cys\text{-}Cys \underset{k_2}{\overset{k_1}{\rightleftharpoons}} SCysAlb + Cys \quad (1)$$

$$2Cys + O_2 \xrightarrow[k_3]{Cu(II)} Cys\text{-}Cys + H_2O_2 \quad (2)$$

Where AlbSH represents the native reduced form of albumin, Cys-Cys represents cystine, SCysAlb stands for S-cysteinylated albumin, and Cys stands for

cysteine[102].

All the reactants listed in the equations above are naturally present in P/S samples. The average initial starting concentrations of albumin, cysteine, cystine, and copper in P/S are all known from previously published literature or simply measured by Jeffs et al., that are 650 ± 60 μM[126], 11 ± 1 μM, 70 ± 10 μM[127, 128] and 18.7 ± 9.2 μM[102], respectively. As determined by Jeffs et al.[102], the starting concentration of S-cysteinylated albumin is usually 20 – 30%. Also, oxygen only enters P/S *ex vivo*, because oxygen carried by hemoglobin in red blood cells is not exposed to P/S *in vivo*. And the hydrogen peroxide (H_2O_2) present in P/S is almost immediately decomposed by catalase present in P/S, with no interference with the S-cysteinylation of albumin. Besides, the rate law for reaction 2 (k_3 = 0.13 M^{-1} s^{-1}) can be estimated from the value for 37 ℃ reported by Kachur et al.[124] And finally, the rate laws for the forward and reverse reactions in reaction 1 at 23 ℃ have been determined by Jeffs et al.[102], with k_1 and k_2 reported to be 0.095 ± 0.017 M^{-1} s^{-1} and 3.37 ± 0.44 M^{-1} s^{-1}, respectively.

As mentioned previously, the Borges group has developed a simple dilute-and-shoot procedure for the ΔS-Cys-Alb assay[102]. This assay meets most of the ideal QA/QC marker requirements, including: targeting an endogenous analyte (albumin in P/S), requiring very low sample volume (10 μL), minimal sample preparation (less than 5 minutes preparation time), inexpensive and rapid, simultaneous change with the most fragile time-sensitive analytes in P/S, and unaffected by any stabilization practices such as protease inhibitors. Moreover, the theoretical reference range for ΔS-Cys-Alb in > 99% of the population has already been estimated as 11-39% from the known population reference ranges for albumin[126], cysteine, and cystine[127, 128]—assuming albumin is the only oxidative consumer of Cys equivalents. Recent findings by Jeffs et

al.[102] in fresh P/S collected from non-acute cardiovascular patients have empirically confirmed this range. (Further validation in fully healthy donors is presently underway.) With the rate laws established for the *ex vivo* formation of S-Cys-Alb and verified to accurately model these reactions in actual P/S[102], it is possible for the ΔS-Cys-Alb assay to effectively place an exposure time stamp on every sample. The application of this ΔS-Cys-Alb assay onto a large set of clinical plasma samples collected under a single SOP but from different collection sites in a multisite clinical trial is described in detail in Chapter 4.

Figures

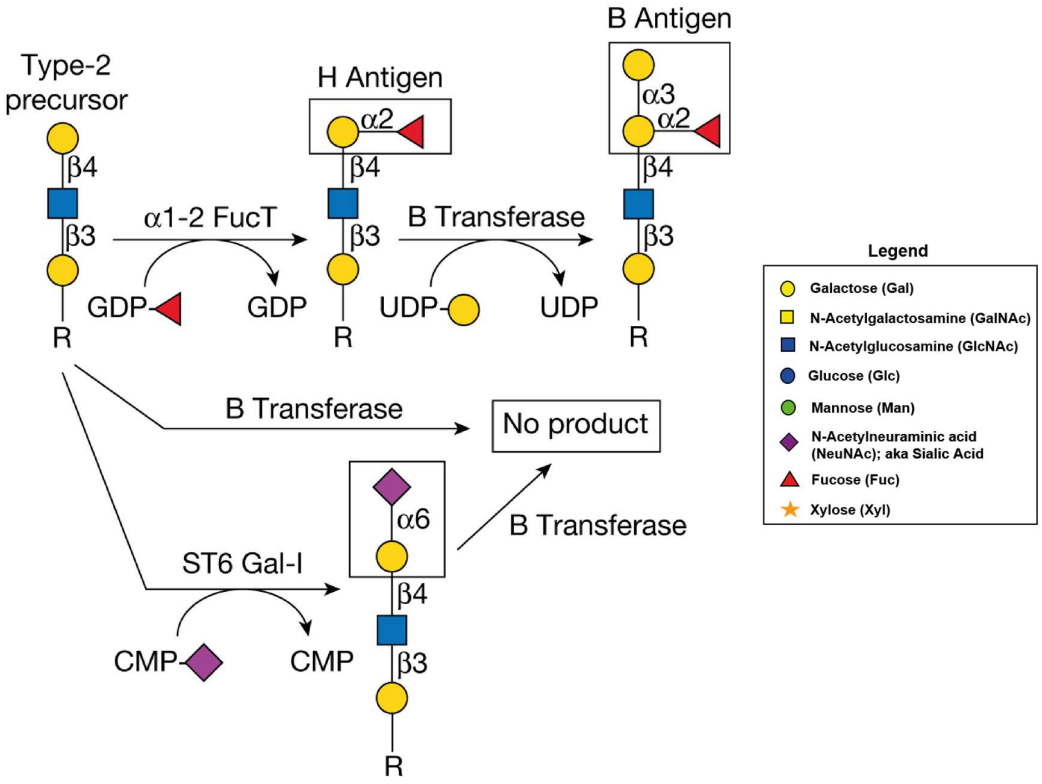


Figure 1.1. The strict acceptor substrate specificity of glycosyltransferases is illustrated by the human B blood group α1-3 galactosyltransferase. The B transferase adds galactose in α1-3-linkage to the H antigen (*top*). This enzyme requires the α1-2-linked fucose modification of the H antigen for activity because the B transferase does not add to an unmodified type-2 precursor (*middle*), or precursors modified by sialyl residues (*bottom*) or other monosaccharides (not shown). The monosaccharide symbol codes are shown in the legend on the right. Adapted with permission from Rini et al.[11].

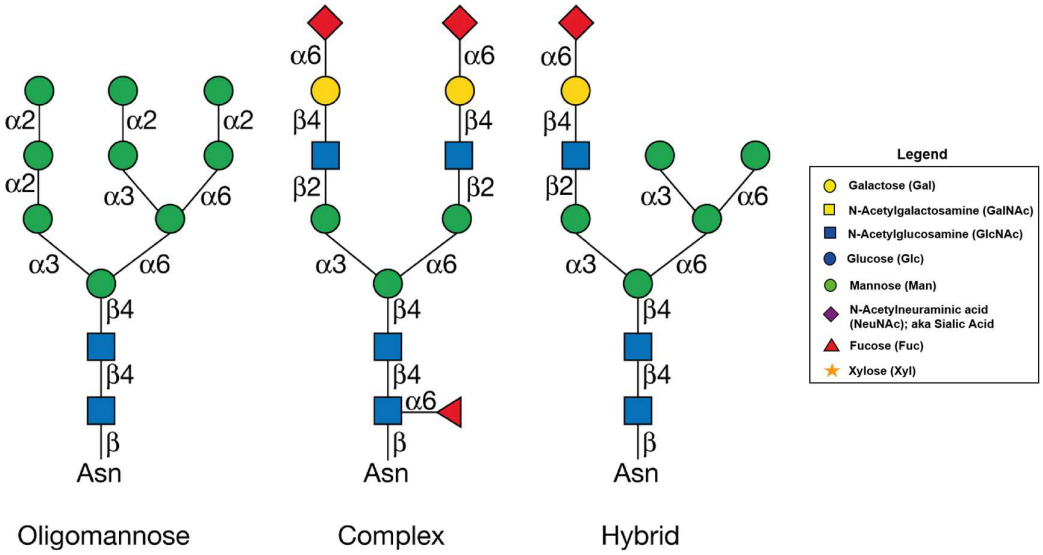


Figure 1.2. Types of N-glycans. N-Glycans at Asn-X-Ser/Thr sequons in eukaryote glycoproteins are of three general types: oligomannose, complex, and hybrid. Each N-glycan contains the common core $Man_3GlcNAc_2Asn$. Adapted with permission from Stanley et al.[129].

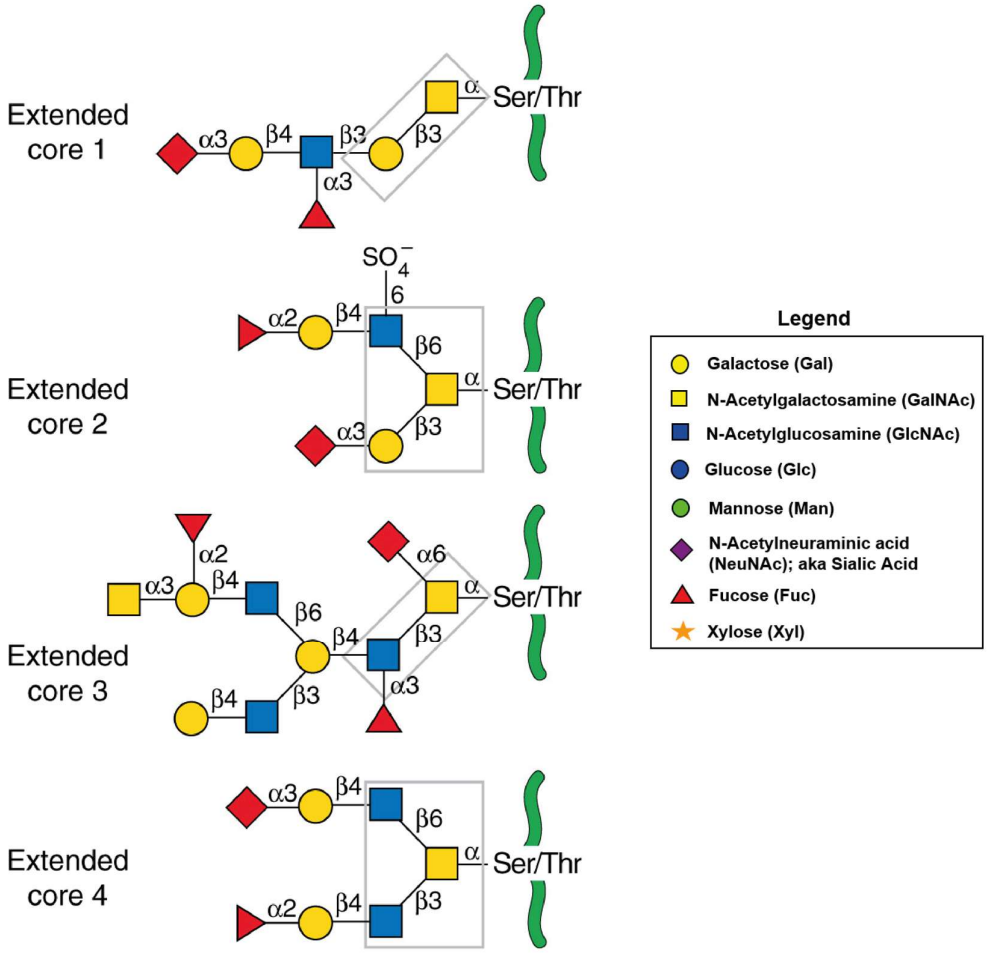


Figure 1.3. Complex O-GalNAc glycans with different cores (identified by gray boxes). Extended core 1, 2, 3, or 4 O-GalNAc glycans are from human respiratory mucins and the extended core 3 O-GalNAc glycan is from human colonic mucins. All four core structures (in *boxes*) may be extended, branched, and terminated with Fuc, Sia, or blood group antigenic determinants. Core 1 and 3 O-GalNAc glycans may also carry α2-6 Sia linked to the core GalNAc. *Green lines* are protein. Adapted with permission from Brockhausen et al.[130].

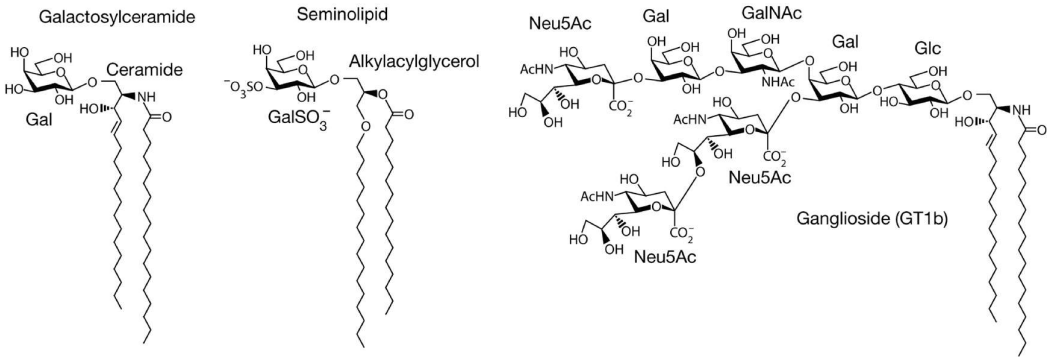


Figure 1.4. Structures of representative glycosphingolipids (GSLs) (left) and glycoglycerolipids (middle), and a complex sialylated GSL (right). GSLs, such as GalCer, are built on a ceramide lipid moiety that consists of a long-chain amino alcohol (sphingosine) in amide linkage to a fatty acid. In comparison, glycoglycerolipids, such as seminolipid, are built on a diacyl or acylalkylglycerol lipid moiety. Most animal glycolipids are GSLs, which have a large and diverse family of glycans attached to ceramide. Shown in right is one example of a complex sialylated GSL, GT1b ($IV^3Neu5AcII^3[Neu5Ac]_2Gg_4Cer$). Adapted with permission from Schnaar et al.[29].

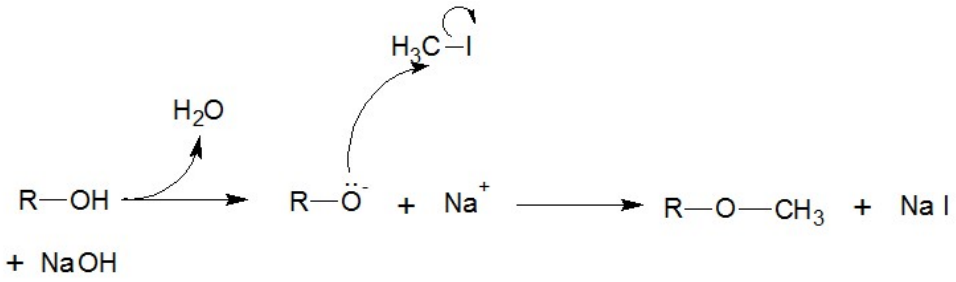


Figure 1.5. Substitution nucleophilic bimolecular (S_N2) mechanism of permethylation step.

CHAPTER 2

DIAGNOSTIC AND PROGNOSTIC PERFORMANCE OF BLOOD PLASMA GLYCAN FEATURES IN THE WOMEN EPIDEMIOLOGY LUNG CANCER (WELCA) STUDY

2.1 Introduction

Lung cancer is the leading cause of cancer death in the United States, which accounts for approximately 25% of cancer deaths[63]. About 33 % and 40% of lung cancer patients are diagnosed at stage IIIB and IV respectively[131], primarily due to a lack of early stage symptoms. The five-year survival rate of stage IV patients is only ~ 5%[63]. Conversely, if lung cancer can be detected before it escapes the lungs, the five-year survival rates usually exceed 50%[63]. Therefore, to improve the outcome of lung cancer patients, a major clinical priority is to detect lung cancer early. Recently, the National Lung Screening Trial (NLST) applied low dose chest computed tomography (LDCT) in older, high-risk individuals and achieved 20% reduction in lung cancer mortality. Yet the positive screening rate in this study was 24.2%, of which 96.4% were false-positive results[132]. The high false-positive rate may lead to additional clinical tests and unnecessary treatments, as well as unnecessary time and costs spent. Thus, a reliable, and highly specific non-invasive blood test could help to reduce the false-positive and overdiagnosis rate of CT scans.

Biomarkers from easily accessible biofluids, such as blood plasma and serum (P/S), could be used as a non-invasive and cost-effective way to improve lung cancer diagnosis and screening. Numerous P/S biomarkers for lung cancer have been extensively studied, including proteins (such as cytokeratin 19 fragments[133, 134] and carcinoembryonic antigen[135, 136]), miRNAs (such as miR-34[137] and miR-182[138, 139]),

methyl-DNA (such as P16[140] and BRMS1[141]), and circulating tumor cells[142]. However, biomarkers with more improved sensitivity and specificity are still in needed.

Aberrant glycosylation is a well-established hallmark of cancer and seems to facilitate the metastasis of various tumor cells[10]. Thus, blood plasma and serum glycomics represents a promising source for a new generation of cancer biomarkers. At present, almost all P/S glycomics studies focus on the analysis of intact glycans, primarily N-linked glycans with O-linked and lipid-linked glycans usually excluded. Generally, a great many intact glycan structures need to be investigated in order to fully capture and quantify the cancer-specific behavior of one unique glycan feature, such as core fucosylation, α2-6 sialylation or β1-4 branching[143]. Glycan node analysis, is a molecularly bottom-up approach to P/S glycomics developed by Borges et al. in 2013, that focuses on monosaccharides and linkage specific "glycan nodes" rather than the intact glycan structures[71, 72, 74, 75, 144]. This approach captures all P/S glycans including N-, O-, and lipid-linked glycans and breaks them down into monosaccharides that maintain linkage information. In short, the method involves the application of glycan linkage (methylation) analysis to whole biofluids (**Fig. 2.1 and 2.2**). Uniquely in this approach, linkage-related glycan features are captured and quantified as single analytical signals, rather than being spread across numerous intact glycans that bear the specific feature. For example, 6-linked galactose and 2,4-linked mannose, corresponding to α2-6 sialylation and β1-4 branching, respectively, are both captured as single chromatographic peak areas (**Fig. 2.1**). In addition, many glycan nodes serve as direct surrogates for the activities of specific glycosyltransferases (GTs)—enzymes that facilitate the construction of glycans.

Recently, we have applied glycan node analysis to several cancer case-control

studies, including pancreatic[71], ovarian[71], prostate[71], bladder[72], breast[144], and lung[71, 74] cancer cohorts. The purpose of this study was to further validate glycan node analysis as a means of detecting and predicting patient outcomes in lung cancer. In addition, though glycan node analysis has been analytically validated in the past[74, 75], we felt it was important to conduct a more comprehensive stability study than that which we have previously reported. The cohort of specimens to which we had access that most readily lent itself to addressing both of these goals was from a study of lung cancer in women. Interestingly, there are several important gender differences in lung cancer, including the facts that 1) after adjusting for the number of cigarettes smoked women have a three-fold greater risk of lung cancer than men[145-148], 2) never-smoker women are at significantly greater risk for lung cancer than men[149], and 3) women tend to have better survival rates than men[150, 151]. As such, we felt that for any differences observed in this study relative to our previously reported results in lung cancer [71], it would also be important to look for any existing gender-based differences in glycan nodes as they may occur in the context of lung cancer.

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 obtained from Cambridge Isotope Laboratories. Acetone was acquired from Avantor Performance Materials (Center Valley, PA). Methanol was purchased from Honeywell Burdick & Jackson (Muskegon, Ml). Acetonitrile and methylene chloride were obtained from Fisher Scientific (Fair Lawn, NJ). DMSO, iodomethane (99%, Cat. No. I8507), chloroform, trifluoroacetic acid (TFA), ammonium hydroxide, sodium borohydride, acetic anhydride, sodium acetate and sodium hydroxide beads (20-40

mesh, Cat. No. 367176) were acquired from Sigma-Aldrich. Pierce spin colunns (900 µL volume) were purchased from ThermoFisher Scientific (Waltham, MA, Cat.No. 69705). GC-MS autosampler vials and Teflon-lined pierceable caps were obtained from ThermoFisher Scientific. GC consumables were acquired from Agilent (Santa Clara, CA); MS consumables were obtained from Waters (Milford, MA).

2.2.2 Plasma/Serum Samples

Women Epidemiology Lung Cancer (WELCA) Set: EDTA plasma samples from stage I-IV lung cancer patients and age-matched controls were collected at 12 different collection centers in France under the supervision of Dr. Isabelle Stücker. This study was approved by the Institutional Review Board of the French National Institute of Health and Medical Research and by the French data Protection Authority (IRB-Inserm, no. 3888 and CNIL no. C13-52). As part of the WELCA Study, all-female lung cancer patients were recruited between September 2014 and November 2016, and age-matched all-female controls were recruited between June 2015 and December 2016. All women living in Paris and the lle de France area, newly diagnosed with lung cancer were considered as eligible cases. Age matched controls were randomly sampled from women living in the same area without a history of cancer. All Peripheral blood samples were drawn and processed following a written standardized protocol[150]. Briefly, after transport to the laboratory at 4 °C, blood samples collected in tubes containing EDTA additive were spun for 15 min at 3000 rpm and 4 °C in a standard centrifuge. Then the collected plasma samples were aliquoted and periodically transported in dry ice to the central repository for final storage at -80 °C. No freeze-thaw cycles occurred prior to shipment to Arizona State University (Borges lab) for analysis. A detailed profile of the clinical characteristics of the patients in this WELCA study is given in **Table 2.1**.

Dual Gender Lung Cancer Set: Sodium heparin plasma samples from a lung cancer study consisting of patients and controls in both genders were collected by Dr. Xifeng Wu at the University of Texas MD Anderson Cancer Center. Even though it is a glycosaminoglycan itself, heparin is featured with monomer units that are predominately carboxylated, sulfated or both, and thus cannot be directly detected by the analytical methodology used in this study. As we have reported previously, there are only negligible differences between glycan nodes measure in heparin plasma vs. EDTA plasma or serum[71], and thus direct comparisons were made for these three types of biospecimens. Venous blood samples were collected from newly diagnosed and histologically confirmed lung cancer patients prior to therapy at the MD Anderson Cancer Center hospital. And the blood samples of age-, gender-, and ethnicity-matched controls were collected at the Kelsey-Seybold Clinic. All blood samples were collected since 1995 and processed following the same SOP. These specimens has previously been described[71].

Stage I Only Lung Cancer Set (also Dual Gender): Serum samples for dual gender stage I lung adenocarcinoma patients were collected together with age-, gender-, and smoking-status-matched controls, under NYU IRB approval at the NYU Langone Medical Center by Dr. Harvey Pass. Arterial blood samples were drawn from fasting patients undergoing surgery between September 2006 to August 2013 to remove one or more lung nodules that were detected during a CT scan. A pathological exam of the excised nodules was performed to determine whether nodules were benign or malignant. Serum was collected under a standardized procedure. These specimens have previously been described[71].

Additional Biospecimen Details: A summary of the case-control sample sets

discussed in this study is provided in **Table 2.2**. A 300-mL plasma sample from an individual donor was obtained from Bioreclamation, which served as a quality control sample to ensure batch-to-batch quantitative reproducibility. All specimens were stored at -80 °C prior to analysis.

2.2.3 Experimental Procedures

The glycan node analysis procedure was adapted from Borges et al[74, 75].

Permethylation, Nonreductive Release, and Purification of Glycans: Nine microliters (9 μL) of blood plasma and 1 μL of a 5 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 were mixed in a 1.5 mL Eppendorf tube, followed by the addition of 270 μL of DMSO. About 0.7 g sodium hydroxide beads were collected in a Pierce spin column (900 μL volume) and washed once with 350 μL of acetonitrile (ACN) followed by two rinses with 350 μL of DMSO. The plasma mixed in DMSO solution was added to the pre-conditioned NaOH beads in the plugged microfuge spin column, followed by adding 105 μL of iodomethane and immediate stirring. After occasional gentle stirring the sample solution in NaOH column for 11 min, the microfuge spin column was unplugged and spun for 30 s at 5000 rpm (1000 g in a fixed-angle rotor). The collected sample solution was quickly transferred into 3.5 mL of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer (pH 7) within a silanized 13 × 100 glass test tube. To maximize glycan recovery, the NaOH beads were then washed twice by 300 μL of ACN, with all spin-throughs immediately transferred into the same silanized glass test tube. To perform liquid/liquid extraction, 1.2 mL of chloroform was added to each test tube, which was then capped and shaken well. After brief centrifugation (as above), the aqueous layer (top layer) was discarded and then replaced by a fresh aliquot of 3.5 mL

of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer (pH 7). With three L/L extractions rounds, the chloroform layer was finally extracted and dried under a gentle stream of nitrogen at 74 °C.

Hydrolysis, reduction, and acetylation: To perform TFA hydrolysis, each sample was mixed with 2 M TFA (325 μL) and incubated at 121 °C for 2 h, which was then dried under a gentle stream of nitrogen at 74 °C. The next step is reduction of sugar aldehydes, in which each sample was incubated at room temperature for 1 h after dissolution in 475 μL of freshly made 10 mg/mL sodium borohydride in 1 M ammonium hydroxide. To remove excess borate, to each sample 63 μL of methanol (MeOH) was added and dried under nitrogen, followed by adding 125 μL of 9:1 (v/v) MeOH : acetic acid. Samples were then dried under nitrogen and fully dried in a vacuum desiccator for 20 min. The last step is acetylation of nascent hydroxyl groups, in which 18 μL of deionized water was added to each test tube, thoroughly mixed and sonicated for 2 min to dissolve any precipitates. After adding 250 μL of acetic anhydride, each sample was incubated for 10 min at 60 °C, followed by mixing with 230 μL of concentrated TFA and incubated again at 60 °C for 10 min. To clean up the sample mixture, L/L extraction was performed twice after adding 1.8 mL of dichloromethane and 2 mL of deionized water to each test tube. With the aqueous layer (top layer) discarded for each round, the organic layer of each sample was then transferred to a silanized autosampler vial, dried under nitrogen and reconstituted in 120 μL of acetone, which was then capped in preparation for injection onto the GC-MS.

Gas chromatography-mass spectrometry: An Agilent Model A7890 gas chromatograph (equipped with a CTC PAL autosampler) coupled to a Waters GCT (time-of-flight) mass spectrometer was employed to analyze the prepared samples. For

each sample, 1 μL of the 120 μL total volume was injected onto a hot (280 °C), silanized glass liner (Agilent Cat. No. 5183-4647) containing a small plug of silanized glass wool at a split ratio of 20:1. A 30-m DB-5ms GC column was used to separate different sample components, facilitated by the carrier gas (helium) with a 0.8 mL/min flow rate. The GC oven temperature was initially kept at 165 °C for 0.5 min, then increased to 265 °C at a rate of 10 °C/min, followed by immediate ramping to 325 °C at a rate of 30 °C/min and finally held at 325 °C for 3 min. Sample components eluted from GC column were subjected to electron ionization (70 eV, 250 °C). Positive-ion mode mass spectra from individual TOF pulses over a *m/z* range of 40-800 were summed every 0.1s. Daily tuning and calibration of the mass spectrometer was performed with perfluorotributylamine to ensure reproducible relative abundances of EI ions and mass accuracy within 10 ppm.

2.2.4 Data Analysis

Quanlynx 4.1 software was employed to integrate the summed extracted-ion chromatogram (XIC) peak areas for all glycan nodes. The peak areas were automatically integrated and manually verified, then exported to a spreadsheet for further analysis.

Two possible normalization approaches were considered: 1) individual hexoses were normalized to heavy glucose, and individual N-acetylhexosamines (HexNAcs) were normalized to heavy N-acetyl glucosamine (GlcNAc); 2) individual hexoses were normalized to the sum of all endogenous hexoses, and individual HexNAcs were normalized to the sum of all endogenous HexNAcs. The second normalization approach tends to provide better inter-batch reproducibility (< 9% for the six most elevated glycan nodes), but the first approach performs better in identifying the potential

increases of all glycan nodes in the patient groups relative to the control group while still keeps acceptable average inter-batch %CV (i.e., < 21%). Thus, results reported below are based on normalization with heavy glucose and heavy GlcNAc, unless otherwise stated.

For the glycan node data of each cohort, outliers were removed by log-transformation and the ROUT method at Q = 1% using GraphPad Prism 7. Outlier-removed data were then reverse transformed by taking the anti-log of each value. To identify differences between cohorts, the Kruskal-Wallis test followed by the Benjamini-Hochberg false discovery correction procedure was performed at a 5% false discovery rate using GraphPad Prism 7. RStudio Version 1.0.143 was used to compare different receiver operating characteristic (ROC) curves by Delong's test or Bootstrap test. The ROC curves shown in figures were plotted by GraphPad Prism 7. Correlation of glycan nodes with age or smoking pack years were assessed via spearman's rank correlation in GraphPad Prism 7. Stage-by-stage multivariate modeling was performed using multivariate logistic regression in RStudio Version 1.0.143, with assessment carried out by leave-one-out-validation, and model selection done using a best subsets procedure. The ability of specific glycan nodes to predict lung cancer survival was evaluated with Cox proportional hazards regression model in SAS 9.4. And GraphPad Prism 7 was applied to generate survival curves and perform associated log-rank Mantel-Cox tests.

2.3 Results

2.3.1 Altered Glycan Features in Stage I-IV Patients

Basic clinical characteristics and n-values of the WELCA sample set were described in the Methods section and **Table 2.1**. All 207 control and 208 stage I-IV

patient samples were randomized and analyzed in 27 batches. Within each control and case sample, a total of 19 glycan "nodes" were measured. The relative abundances of each of these nodes contributed at least 1% of the total hexose or total N-acetylhexosamine (HexNAc) signal. Data from each of the 19 glycan nodes were normalized to heavy, isotope-labeled glucose and GlcNAc internal standards. Statistically significant differences were detected in each cancer stage relative to the control cohort: 10, 6, 18 and 19 out of 19 glycan nodes were increased in stage I, II, III and IV respectively (**Table 2.3**). Data for each glycan node normalized to the sum of endogenous hexoses or HexNAcs were analyzed analogously (**Table 2.4**). This revealed shifts in glycan compositions in stage I-IV patients vs. controls. However, because quantitative changes in glycans tended to outpace glycan compositional changes (as we have previously observed[71]) this normalization procedure was not as sensitive in distinguishing age-matched controls from lung cancer patients at each stage.

Six glycan nodes were found to be significantly elevated at nearly every stage in lung cancer patients relative to the age-matched controls, these included: 2-linked mannose and 4-linked N-acetylglucosamine (GlcNAc), both of which are mostly associated with total glycosylation levels especially for N-glycans[129]; 6-linked galactose, corresponding to α2-6 Sialylation[74]; 2,4-linked mannose, corresponding to β1-4 Branching[74]; 2,6-linked mannose, corresponding to β1-6 Branching[74]; and 3,4-linked GlcNAc, which primarily corresponds to outer-arm fucosylation[74] (**Fig. 2.3**). The latter four nodes were amongst the top five most elevated nodes in our previously reported lung cancer study[71]. The receiver operating characteristic (ROC) curve c-statistics (areas under the curve, AUCs) for these six glycan nodes in stage I-IV patients vs. controls ranged (with two exceptions) from 0.68 to 0.92 (**Fig. 2.3**).

2.3.2 Prominent Early-Stage Alteration

Relative to the age-matched controls, five of the six top performing glycan node markers in stage I patients, and four in stage II patients, were significantly increased (**Fig. 2.3**). In addition, the ROC c-statistics (AUC) of these glycan nodes were mostly statistically significant and ranged from 0.68 to 0.80 (with one exception). The notable alterations of glycan nodes in early stages were usually not observed for other lung cancer sets, such as the dual gender lung cancer set and stage I only lung cancer set (which was also dual-gender) reported in our previous work[71] (n-values for these studies provided in **Table 2.2**). ROC curves of these three lung cancer sets were compared (**Table 2.5**) and shown in **Fig. 2.4**. Significant differences were observed for β1-6 branching when the ROC curve of the stage I cohort of the WELCA set was compared to that of the stage I only lung cancer set and that of the stage I cohort of the dual gender lung cancer set. When comparing ROC curves for the stage IV cohorts of the WELCA set and the dual gender lung cancer set, significant differences were found for three glycan features--i.e., α2-6 sialylation, β1-4 and β1-6 branching. Since all the lung cancer patients and age-matched controls involved in the WELCA set were female, the gender dependence of these glycan node markers in early stages was evaluated in the other two lung cancer sets, which included patients and controls from both sexes. When sample set and stage were held constant, the ROC curves of two sexes were compared for each glycan node using Delong's test or the Bootstrap test (**Table 2.6**). No significant differences were observed, however, indicating the early-stage clinical performance characteristics of the six glycan node markers were independent of gender.

2.3.3 Negligible dependence on smoking-status, age and histological type

No significant alteration of five out of the six top performing glycan node

markers was observed in a direct comparison of never-smokers, previous smokers and current smokers within the control cohort. The only exception was 3,4-linked GlcNAc (corresponding to outer-arm fucosylation), which was slightly elevated in current smokers relative to previous smokers (**Fig. 2.5**). Spearman's rank correlation analysis demonstrated no statistically significant correlation with smoking pack-years in the control cohort. Together, these data revealed that the top performing glycan node markers within the control cohort had negligible dependence on smoking status. (A parallel analysis within the cancer patient cohort was not conducted due to the confounding correlation between smoking and lung cancer.)

The average ages of the control and case cohorts were nearly identical (61.2 and 61.6, respectively; **Table 2.1**). After pooling all data from the cases (all stages) and controls, 3,4-linked GlcNAc, corresponding to outer-arm fucosylation, was found to be weakly correlated with age (correlation coefficient r = 0.159, Bonferroni-corrected p = 0.0096; **Fig. 2.6a**). No statistically significant correlations with age were observed for the other five top performing glycan nodes (**Table 2.7**). When the population was divided into smaller age groups, only 3,4-linked GlcNAc showed significant differences between pairs of decades; if the control and case cohorts were investigated in isolation, 3,4-linked GlcNAc within the controls in particular indicated a distinct upward pattern in more advanced age groups (**Fig. 2.6b**). The same phenomenon was observed in the male-only controls of the dual gender lung cancer set. However, with the exception of 3,4-GlcNAc, these analyses indicated a general lack of dependence of glycan node markers on age.

The effect of lung cancer histological subtypes on the six glycan nodes was evaluated in the stage IV non-small cell lung cancer (NSCLC) subcohort (i.e., the

largest single-stage subcohort available; **Table 2.1**). For each glycan node marker, ROC curves of the three histological subtypes of NSCLC – adenocarcinoma, squamous cell carcinoma, and large cell carcinoma – were compared pairwise by Delong's test or Bootstrap test (**Table 2.8**). No statistically significant differences between histological subtypes of NSCLC were discovered for any glycan node marker.

These findings on glycan node independence from smoking status, age and histological type are consistent with our previously reported findings from other lung cancer case/control studies[71].

2.3.4 Total Glycosylation and Multivariate Model of glycan features

The clinical performance characteristics of total glycosylation (i.e., total hexoses, total HexNAcs, and the sum of total hexoses & total HexNAcs) were evaluated and compared to individual glycan node markers on a stage-by-stage basis (**Table 2.9**). Results of ROC curve comparisons by paired Delong's tests demonstrated that total glycosylation cannot distinguish stage I-IV cases from controls better than individual glycan node markers.

Additionally, multivariate logistic regression models were built and compared with the clinical performance characteristics of individual glycan nodes at each stage (**Fig. 2.7**). Fully cross-validated multivariate logistic regression models were no better at detecting lung cancer than the top-performing individual glycan node at each respective stage. Again, these results were consistent with our previous observations in lung cancer[71].

2.3.5 Prediction of All-Cause Mortality

To evaluate the ability of the six glycan nodes to predict all-cause mortality,

glycan node data were broken into quartiles and analyzed by Cox proportional hazards regression, with adjustment for age, smoking status, and cancer stage (**Table 2.10**). First and foremost, for patients in all four stages, the top quartiles of all six glycan node markers predicted all-cause mortality with hazard ratios in the range of 2-3 and $p < 0.01$, relative to all other quartiles combined. The different rates of death for the top quartile versus all other quartiles for each glycan node marker are illustrated by survival curves (**Fig. 2.8**).

When focusing on stage III and IV patients, the top quartiles of all six glycan node markers predicted all-cause mortality with hazard ratios in the range of 2-3 and $p < 0.05$ **(Table 2.10**) relative to all other quartiles combined (survival curves shown in **Fig. 2.9**). Similar results were observed for stage IV patients only (**Table 2.10**). However, when stage III patients were analyzed alone, the hazard ratios of all six glycan nodes were not significantly different from 1 ($p > 0.05$), indicating the relative risk of death was not detectably different between patients in the top quartile vs. all other quartiles of each glycan node. 6-linked galactose (corresponding to α2-6 Sialylation) and 2,4-linked mannose (corresponding to β1-4 Branching), were significantly different between stages III and IV (**Fig. 2.3 b,c**).

Overall, these results for glycan node-based prediction of mortality vary slightly, but are largely consistent with our previously reported results on the ability of α2-6 sialylation and branched mannose residues to predict all-cause mortality in lung cancer.

2.4 Discussion

Six out of 19 quantified glycan nodes, corresponding to total glycosylation levels (especially for N-glycans), α2-6 Sialylation, β1-4 Branching, β1-6 Branching and outer-arm fucosylation, were significantly elevated in the WELCA lung cancer patients

relative to age-matched controls. Most of these findings in the WELCA set are highly consistent with our previously reported lung cancer study on a dual gender lung cancer set[71], which also demonstrated the distinct increase of the latter four glycan features within stage III-IV cases compared to the control cohort.

Our observations of the glycan node-based features changes in lung cancer patients are closely aligned with the intact glycan changes reported by Vasseur and colleagues[69]. Their intact glycan analysis results primarily revealed the significant increase of outer-arm fucosylation, as well as fucosylated tri- and tetra-antennary N-glycans, in line with β1-4 branching, β1-6 branching, in lung cancer samples.

Unlike the other two lung cancer sets that we have reported on previously[71], some glycan node-based features were substantially altered in the WELCA lung cancer patients at stages I-II (**Fig. 2.4**). Even though a relatively low number of samples were measured for early stages (n = 16 and 13 for stage I and II, respectively), statistically significant elevations were detected in most of the six glycan node markers, alongside comparatively high ROC c-statistics. There are two possible causes for this phenomenon. First, since the lung cancer patients and controls enrolled in the WELCA study are all female, a distinct gender dependence of glycan features may exist, especially in early stages. However, this possibility is not evidenced by the observation that no significant difference was detected between men and women in stage I and II of the dual gender lung cancer set, as well as in the stage I only lung cancer set—which was also dual gender (**Table 2.6**). The second possible explanation is that the non-smoking-matched controls of the WELCA set may have much lower relative abundances of all the glycan nodes of interest relative to the smoking-matched controls from other lung cancer sets. In the WELCA set most controls were never-smokers, but

case patients are mainly current-smokers (**Table 2.1**), indicating that the smoking history possibly contributes to the increase of glycan nodes. Taken together the observations of the WELCA set that the top performing glycan node markers within the control cohort had negligible dependence on smoking status (**Fig. 2.6**), smoking appears to contribute to slight, but statistically insignificant elevation of glycan nodes.

The six top performing glycan nodes-based features in this study were not only able to distinguish lung cancer patients from age-matched controls, but were also able to predict all-cause mortality in the WELCA set—a finding that agrees well with our previously reported study on the dual gender lung cancer set[71]. Similar discoveries regarding the prognostic capacity of P/S glycans have also been reported by other groups. Hashimoto and colleagues[152] suggested that specific glycoforms of serum α_1-acid glycoprotein (AGP) seemed to appropriately predict progression and mortality of several carcinomas, including lung cancer. According to their follow-up studies, patients who had the AGP glycoforms that contained highly fucosylated and branched sugar chains tended to have a poor prognosis. Besides the glycan features discussed above, another good prognostic predictor of lung cancer is the sialyl Lewis X epitope (SLe^x)[153], which consists of α2-3 sialylation instead of α2-6 sialylation. The progression and survival in non-small-cell[154-156] and small-cell lung cancer[157] can both be predicted by SLe^x.

Many studies have reported important gender differences in lung cancer between men and women, in terms of histological type, tobacco exposure, survival and treatment response[158, 159]. Here, by comparison with our previously conducted studies[71] , no obvious gender differences were detected with regard to P/S glycan features. Smoking is the primary risk factor for lung cancer. However, a large percentage of

women with lung adenocarcinoma, between 20% and 30% in Western countries and nearly 80% in Asian countries, are non-smokers[150]. Hence, some female-specific risk factors for lung cancer exist and may play vital roles in lung cancer development, progression and survival; these may include hormonal factors and occupational risk factors in female occupations—as suggested by Stucker et al[150]. Therefore, studies focused on lung cancer in women, especially on the gender specific risk factors, should garner further attention as they promise to disentangle the etiology of lung cancer in women.

2.5 Conclusion

The diagnostic and prognostic capacity of plasma glycan features in stage I-IV lung cancer—as represented by monosaccharide and linkage-specific glycan nodes—were validated in the WELCA case-control study. Significant elevation of α2-6 sialylation, β1-4 branching, β1-6 branching, outer-arm fucosylation and total N-glycosylation level, were observed in almost every stage of lung cancer relative to age-matched control groups. Alteration of glycan features in lung cancer was independent of smoking status, age and histological subtypes of lung cancer. Furthermore, the above six glycan features predicted all-cause mortality in lung cancer patients after adjusting for age, smoking status, and cancer stage. Early-stage detection was stronger than we have previously observed[71], but this observation may have been related to the lack of smoking status-matching between cases and controls in the WELCA study. No gender-based differences were discovered in glycan features associated with lung cancer.

Figures

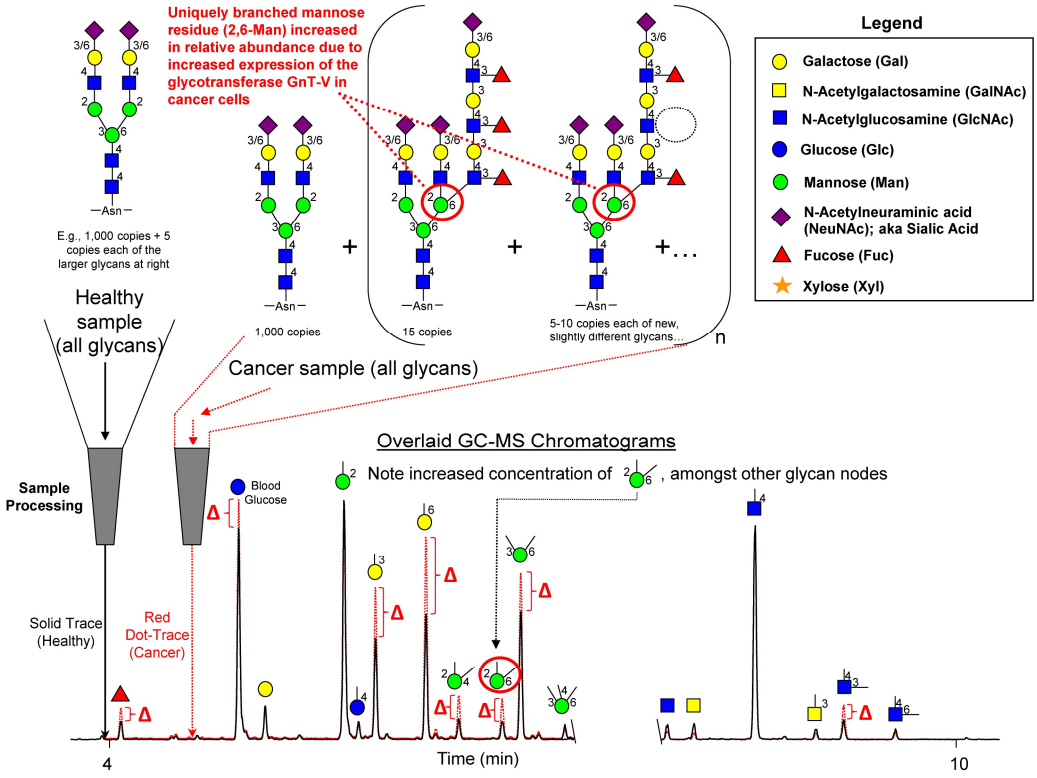


Figure 2.1. Conceptual overview of the glycan "node" analysis concept, which essentially consists of applying glycan linkage (methylation) analysis to whole biofluids. Intact normal and abnormal glycans including O-glycans, N-glycans, and glycolipids are processed and transformed into partially methylated alditol acetates (PMAAs, Figure 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 among all the aberrant intact glycan structures provides a moredirect 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. This figure was adapted with permission from Borges et al.[74]. Copyright 2013 American Chemical Society.

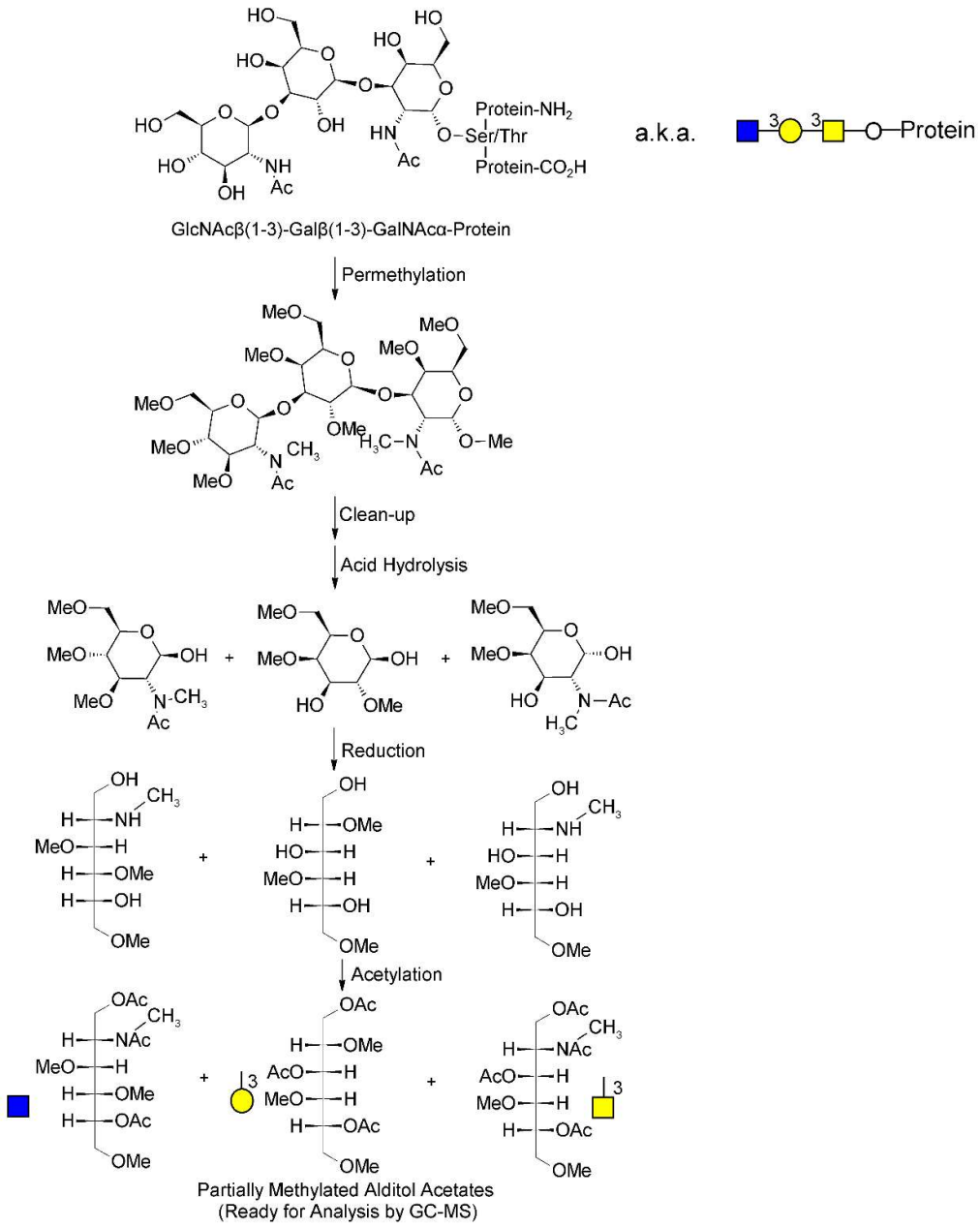
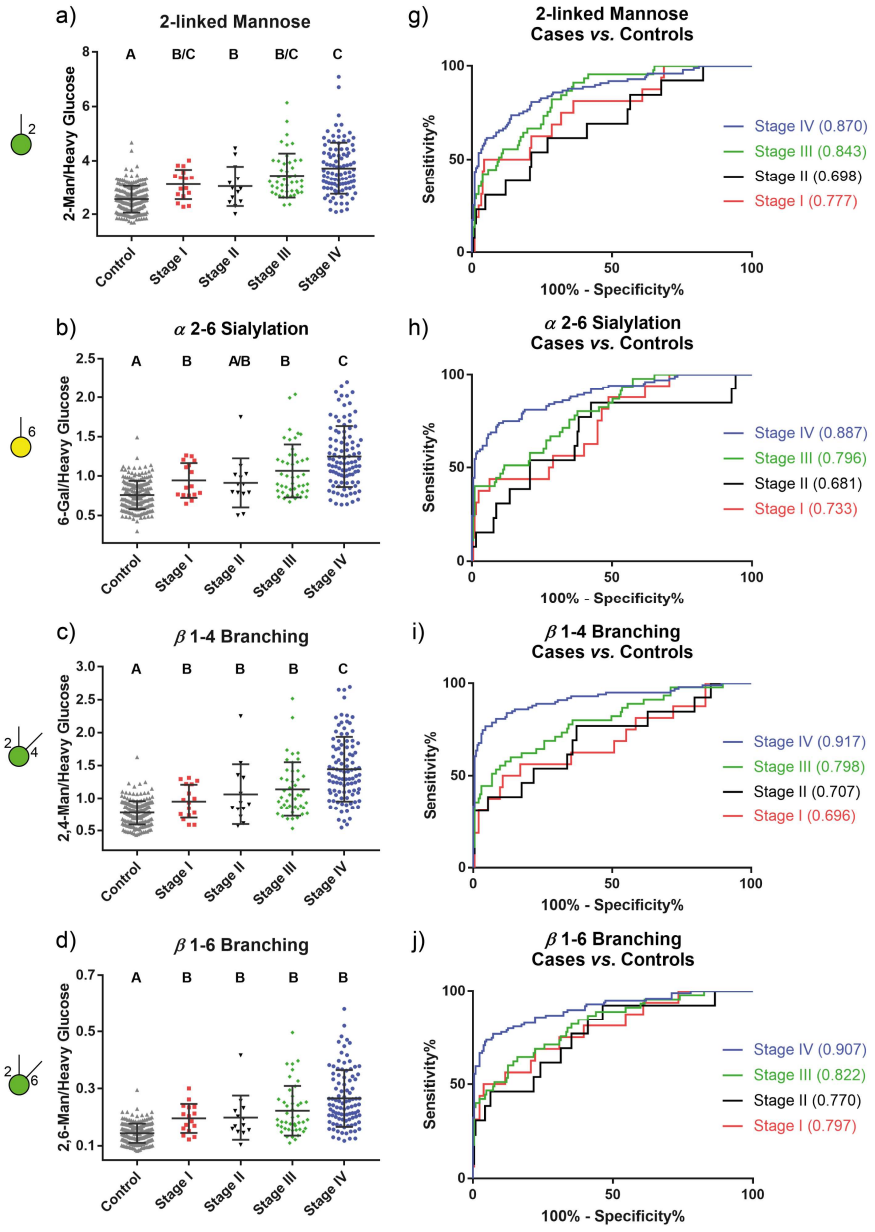


Figure 2.2. Molecular overview of the glycan "node" analysis procedure. For glycans from blood plasma and other biofluids, O-linked glycans are released during permethylation, while N-linked glycans and glycolipids are released during acid hydrolysis. The unique pattern of methylation and acetylation 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 et al.[74]. Copyright 2013 American Chemical Society.


a)
2-linked Mannose
2-Man/Heavy Glucose
A
B/C
B
B/C
C
Control
Stage I
Stage II
Stage III
Stage IV
b)
α 2-6 Sialylation
6-Gal/Heavy Glucose
A
B
A/B
B
C
c)
β 1-4 Branching
2,4-Man/Heavy Glucose
A
B
B
B
C
d)
β 1-6 Branching
2,6-Man/Heavy Glucose
A
B
B
B
B
g)
2-linked Mannose
Cases vs. Controls
Sensitivity%
100% - Specificity%
Stage IV (0.870)
Stage III (0.843)
Stage II (0.698)
Stage I (0.777)
h)
α 2-6 Sialylation
Cases vs. Controls
Stage IV (0.887)
Stage III (0.796)
Stage II (0.681)
Stage I (0.733)
i)
β 1-4 Branching
Cases vs. Controls
Stage IV (0.917)
Stage III (0.798)
Stage II (0.707)
Stage I (0.696)
j)
β 1-6 Branching
Cases vs. Controls
Stage IV (0.907)
Stage III (0.822)
Stage II (0.770)
Stage I (0.797)

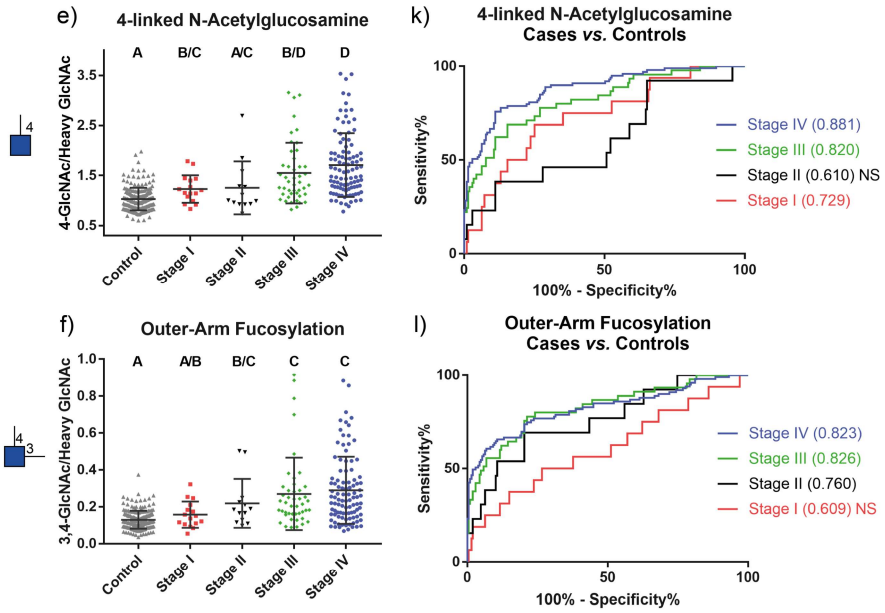


Figure 2.3. Univariate distribution (a-f) and ROC curves (g-l) for the six top performing glycan nodes in the WELCA study. The Kruskal−Wallis test was performed followed by the Benjamini−Hochberg false discovery correction procedure. The letters at the top of data points in panels a-f demonstrate statistically significant differences between groups; any overlap in letters indicates a lack of significant differences between groups. ROC curves for stage I-IV lung cancer cases vs. controls are provided in panels g-l. Areas under the ROC curves are provided in parenthesis next to the specified stages. “NS” next to the AUC values indicates that the ROC curve is not statistically significant. All raw glycan node data in this figure were normalized to heavy glucose or heavy GlcNAc prior to data analysis.

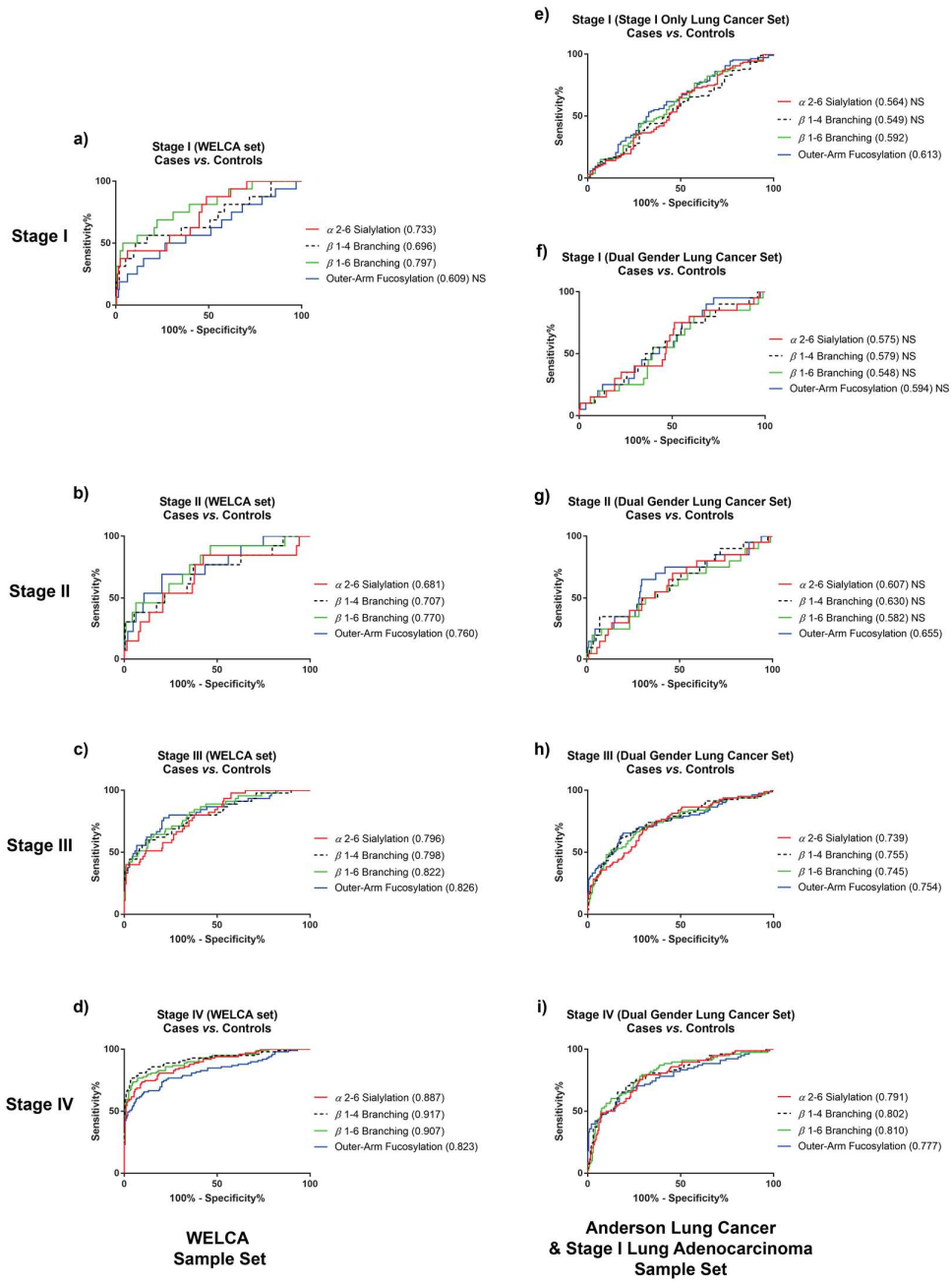


Figure 2.4. ROC curves for four top performing glycan nodes in stage I-IV within different lung cancer sets. Four glycan nodes with highly-ranked performance in all three sample sets were shown. The ROC curves from WELCA sample set are illustrated in panels a-d. In panels e-i are ROC curves from the other two lung cancer sets: Dual Gender Lung Cancer set (e-h) and Stage I Only Lung Cancer set (also dual gender). *N*-values of each group are provided in **Table 2.2**.

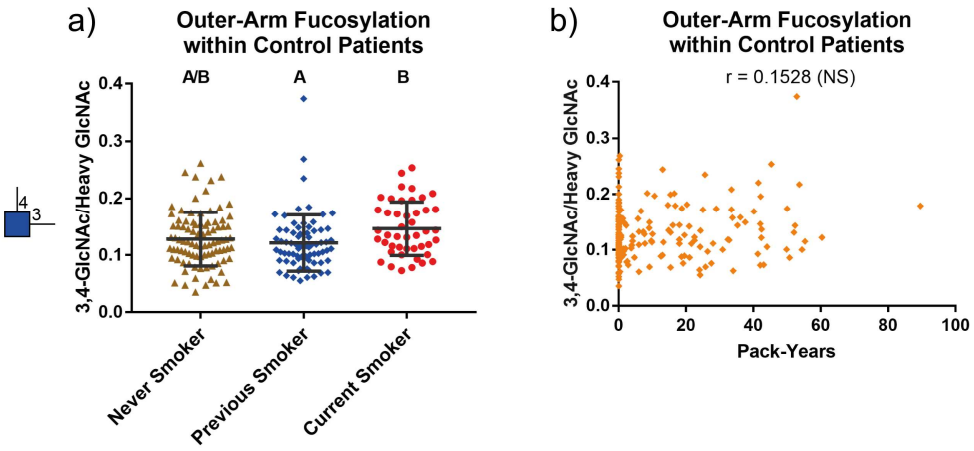


Figure 2.5. Connection between outer-arm fucosylation and smoking status within the control group. (a) The univariate distributions of outer-arm fucosylation within the control group are shown, subdivided by smoking status. Different letters above the data points indicate statistically significant differences between groups as detected by the Kruskal-Wallis test followed by the Benjamini-Hochberg FDR correction procedure. (b) Spearman's rank correlation between outer-arm fucosylation and smoking pack-years. Correlation coefficient is provided above the data points. "NS" next to the correlation coefficient demonstrates a lack of statistical significance. For the other five top performing glycan nodes not shown in this figure, no statistically significant associations with smoking were found.

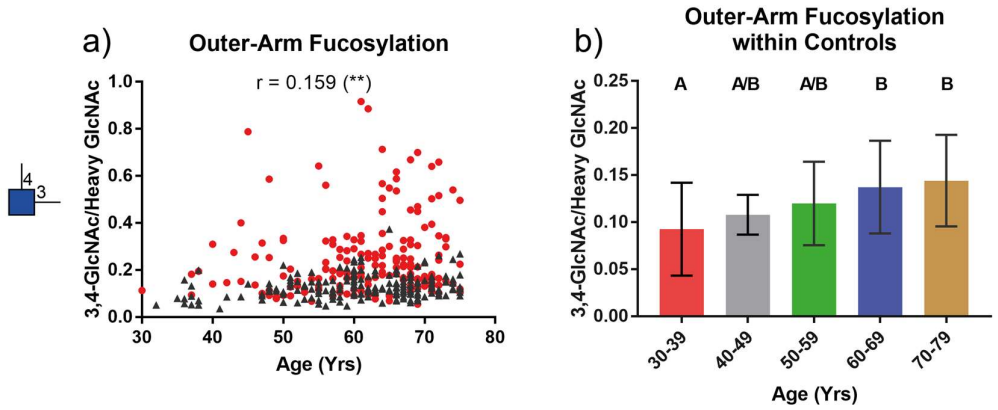


Figure 2.6. Connection between age and a top performing glycan node, 3,4-linked GlcNAc, in the WELCA study. (a) Spearman's rank correlation coefficients are provided above the data points. "**" next to the coefficient indicates that the Spearman's rank correlation was statistically significant with Bonferroni-corrected $p <$ 0.01. Controls are indicated by black triangles and cases by red dots. (b) The Kruskal−Wallis test was performed followed by the Benjamini−Hochberg false discovery correction procedure to identify difference between age groups. The letters at the top of bars demonstrate statistically significant differences between groups; any overlap in letters indicates a lack of significant differences between groups. For the other five top performing glycan nodes not shown in this figure, no statistically significant associations with age were found.

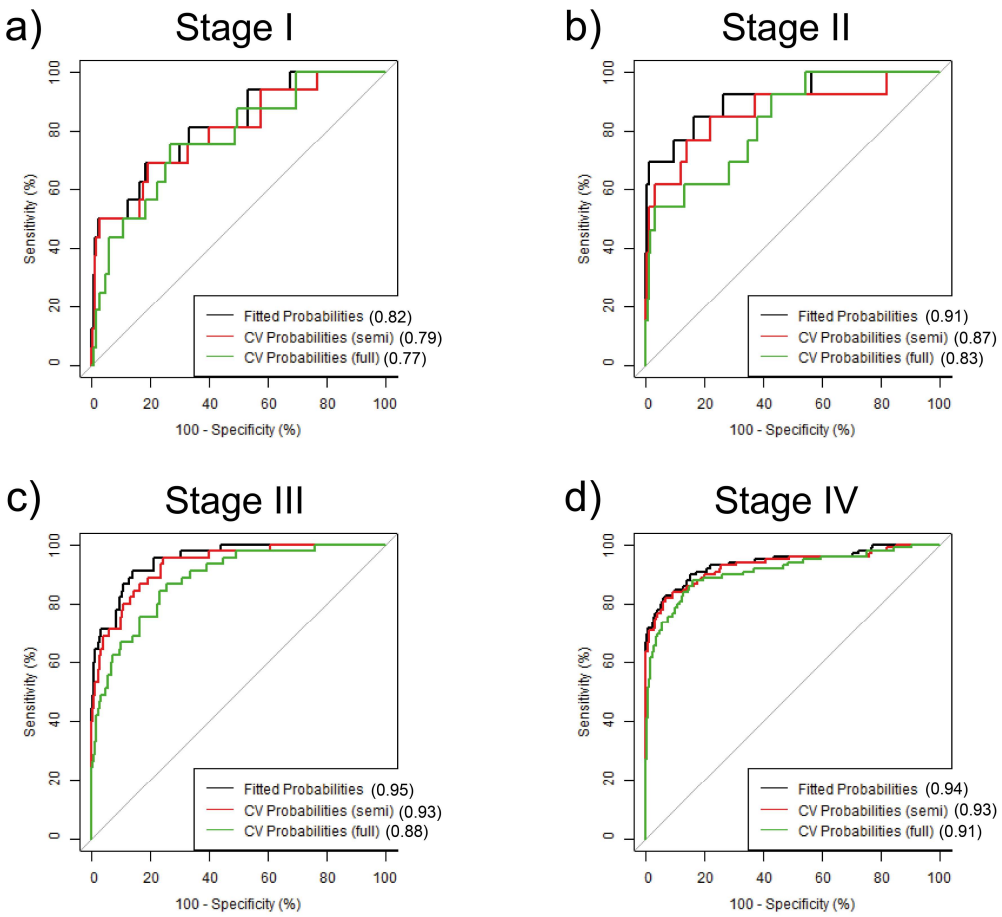


Figure 2.7. Multivariate logistic regression models for stage I–IV patients from the WELCA data set. Three multivariate logistic regression models were built and corresponding ROC curves were plotted for each stage with different fitting procedures: (1) fitted once on the complete data set and acquired probability (referred to as "Fitted Probabilities") with no use of cross-validation; (2) fitted once on the complete data set, cross-validated with fixed predictors but mobile parameter estimates at each iteration (predicted probability referred to as "CV Probabilities (semi)"); and (3) refitted at each iteration of cross-validation (corresponding probability demonstrated as "CV Probabilities (full)"). ROC AUC values are provided in parenthesis next to the specified models. For each stage, the best performing individual glycan node was selected and compared to the fully validated multivariate model. No significant differences were detected (Delong's test).

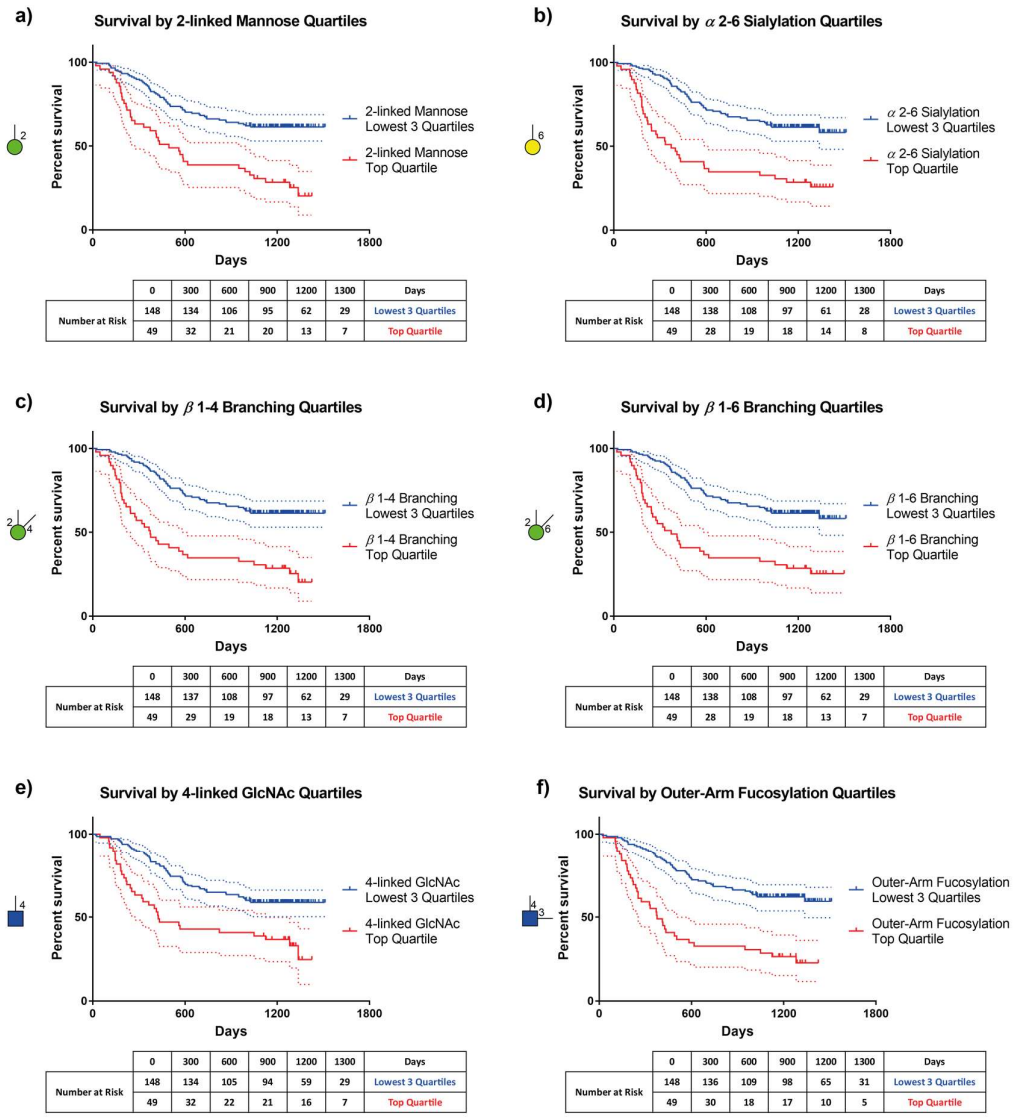


Figure 2.8. Survival curves for the six top performing glycan nodes for all stages combined. In each panel, the top quartile of specified glycan node is compared to all other quartiles combined. According to results of log-rank Mantel-Cox test, the survival curves within each panel are significantly different ($p < 0.0001$). Dotted lines represent 95% confident intervals. The median duration of follow-up for deceased patients (until death) was 406 days; for those that remained alive it was 1253 days. The median follow-up time for all patients was 1057 days.

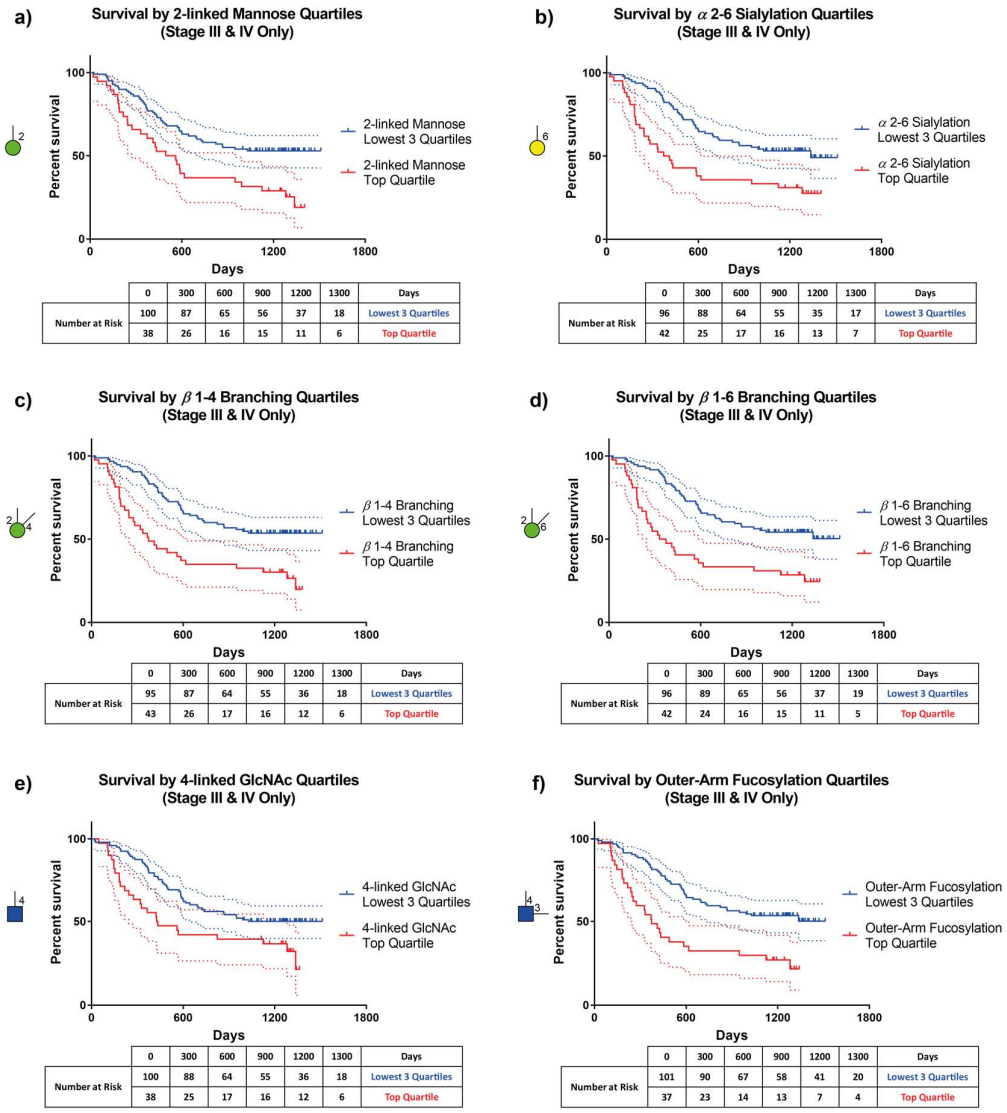


Figure 2.9. Survival curves of the six top performing glycan nodes for stage III and IV combined. In each panel, the top quartile of specified glycan node is compared to all other quartiles combined. According to the results of a log-rank Mantel-Cox test, the survival curves within each panel are significantly different ($p < 0.05$). Dotted lines represent 95% confident intervals. The median duration of follow-up for patients that died, until death, was 393 days; for survivors this value was 1264 days. The median total follow-up time for all patients was 908 days.

Tables

Table 2.1. Basic Clinical Characteristics and N-values of the WELCA Sample Set.

		WELCA set	
		Controls	**Cases**
Age[a]		61.2 ± 9.73	61.6 ± 9.04
Smoking Status	Never-Smoker	90	36
	Previous-Smoker	72	52
	Current-Smoker	45	98
	Unknown	0	22
Staging	Stage I	N/A	16
	Stage II	N/A	13
	Stage III	N/A	45
	Stage IV	N/A	99
	Unknown Stage	N/A	35
Tumor Histological Types	SCLC – Located	N/A	7
	SCLC - Disseminated	N/A	12
	NSCLC - Adenocarcinoma	N/A	131
	NSCLC - Squamous cell carcinoma	N/A	25
	NSCLC - Large cell carcinoma	N/A	13
	NSCLC - Adenosquamous	N/A	1
	NSCLC - Sarcoma	N/A	3
	Unknown	N/A	16

[a]Age in years ± SD.

Table 2.2. Composition of Three Lung Cancer Sample Sets and Their Sub-cohorts.

Name of Sample Set	Plasma or Serum	Controls (M/F)	Stage I (M/F)	Stage II (M/F)	Stage III (M/F)	Stage IV (M/F)
WELCA set	plasma	0/207	0/16	0/13	0/45	0/99
Dual Gender Lung Cancer set	plasma	123/76	14/6	12/8	50/31	47/31
Stage I Only Lung Cancer set	serum	28/45	33/74	–	–	–

Table 2.3. Statistically Significant Differences between Cohorts within the WELCA Study[a]

glycan node[b]	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	iiii	iiii	ns	ns	ns	ns	ns	ns
t-Gal	ii	ns	iiii	iiii	ns	ns	ns	ns	i	ns
2-Man	iii	i	iiii	iiii	ns	ns	ns	ns	i	ns
4-Glc	iii	i	iiii	iiii	ns	ns	ns	ns	ns	ns
3-Gal	ns	ns	ns	iiii	ns	ns	ns	ns	iii	ii
6-Gal	ii	ns	iiii	iiii	ns	ns	i	ns	ii	i
3,4-Gal	ii	ii	iiii	iiii	ns	ns	ns	ns	ns	ns
2,4-Man	i	i	iiii	iiii	ns	ns	iii	ns	ii	ii
2,6-Man	iii	ii	iiii	iiii	ns	ns	ns	ns	ns	ns
3,6-Man	ns	ns	ii	iiii	ns	ns	ns	i	iii	i
3,6-Gal	ns	ns	i	iiii	ns	ns	ns	ns	iii	ii
3,4,6-Man	ii	ns	ii	ii	i	ns	ns	ns	ns	ns
t-GlcNAc	ns	ns	iiii	ii	ns	ns	ns	ii	ns	ns
4-GlcNAc	ii	ns	iiii	iiii	ns	ns	i	i	ii	ns
3-GlcNAc	ns	ns	iiii	iiii	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	ii	iiii	ns	ns	ns	ns	ns	ns
3,4-GlcNAc	ns	ii	iiii	iiii	ns	i	ii	ns	ns	ns
4,6-GlcNAc	iii	ns	iiii	iiii	ii	ns	ns	ii	ii	ns
3,6-GalNAc	ns	ns	ii	iiii	ns	ns	ns	ns	ns	ns

[a]Heavy, stable isotope labeled glucose (Glc) and GlcNAc were utilized to normalize Hexose and HexNAc data, correspondingly.

[b]Kruskal-Wallis test followed by Benjamini−Hochberg false discovery correction procedure at 95% confidence level is given. “ns” stands for “not significant”. “i” indicates $p < 0.05$. “ii” indicates $p < 0.01$. “iii” indicates $p < 0.001$, and “iiii” indicates $p < 0.0001$. C: controls; I: Stage I; II: Stage II; III: Stage III; IV: Stage IV.

Table 2.4. Statistically Significant Differences between Cohorts within the WELCA Study (by Endogenous Normalization [a]).

Glycan Node[b]	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	dd	ns	ns	ns	ns
2-Man	ns	ns	ns	dd	ns	ns	ns	ns	d	d
4-Glc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-Gal	ns	dd	dd	ns	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	i	iiii	ns	ns	ii	ns	ns	ns
3,4-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
2,4-Man	ns	ns	ii	iiii	ns	ns	iiii	ns	ns	iiii
2,6-Man	ns	ii	iii	iiii	ns	ns	ns	ns	ns	ns
3,6-Man	ns	dd	ns	ns	dd	ns	ns	ns	ii	ns
3,6-Gal	ns	dd	ns	ns	ns	ns	ns	ns	ii	ns
3,4,6-Man	ns	dd	ddd	dddd	d	ns	ddd	ns	ns	dd
t-GlcNAc	ns	d	dd	dddd	ns	ns	ddd	ns	ns	ddd
4-GlcNAc	ns	ns	ns	iiii	ns	ns	ns	ns	ns	ns
3-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	dddd	dddd	ns	ns	dd	ns	ns	ns
3,4-GlcNAc	ns	ii	iiii	iiii	i	i	ii	ns	ns	ns
4,6-GlcNAc	ns	d	dd	dddd	dd	dd	dddd	ns	ns	d
3,6-GalNAc	ns	ns	dd	dddd	ns	ns	ns	ns	ns	ns

[a]Hexose data were normalized to the sum of endogenous hexoses, and HexNAc data were normalized to the sum of endogenous HexNAcs.

[b]Kruskal-Wallis test followed by Benjamini–Hochberg false discovery correction procedure at 95% confidence level is given. "ns" stands for "not significant". "i" and "d" stands for "increased" and "decreased". i/d indicates $p < 0.05$. ii/dd indicates $p < 0.01$. iii/ddd indicates $p < 0.001$, and iiii/dddd indicates $p < 0.0001$. C: controls; I: Stage I; II: Stage II; III: Stage III; IV: Stage IV.

Table 2.5. Stage-by-stage ROC Comparison of the Top Performing Glycan Nodes.

Stages[a]	Glycan Feature	A: ROC AUC of set A B: ROC AUC of set B	*p*-value of Delong's test for two ROC curves[b]
Stage I **Set A: WELCA Set** **Set B: Stage I Only Lung Cancer Set**	α2-6 Sialylation	A: 0.733 B: 0.564	0.031 (NS)
	β1-4 Branching	A: 0.696 B: 0.549	0.112 (NS)
	β1-6 Branching	A: 0.797 B: 0.592	0.008
	Outer-Arm Fucosylation	A: 0.609 B: 0.613	0.965 (NS)
Stage I **Set A: WELCA Set** **Set B: Dual Gender Lung Cancer Set**	α2-6 Sialylation	A: 0.733 B: 0.575	0.092 (NS)
	β1-4 Branching	A: 0.696 B: 0.579	0.264 (NS)
	β1-6 Branching	A: 0.797 B: 0.547	0.008
	Outer-Arm Fucosylation	A: 0.609 B: 0.594	0.885 (NS)
Stage II **Set A: WELCA Set** **Set B: Dual Gender Lung Cancer Set**	α2-6 Sialylation	A: 0.681 B: 0.607	0.509 (NS)
	β1-4 Branching	A: 0.707 B: 0.630	0.489 (NS)
	β1-6 Branching	A: 0.770 B: 0.582	0.071 (NS)
	Outer-Arm Fucosylation	A: 0.760 B: 0.655	0.302 (NS)
Stage III **Set A: WELCA Set** **Set B: Dual Gender Lung Cancer Set**	α2-6 Sialylation	A: 0.796 B: 0.739	0.241 (NS)
	β1-4 Branching	A: 0.798 B: 0.755	0.407 (NS)
	β1-6 Branching	A: 0.822 B: 0.745	0.119 (NS)
	Outer-Arm Fucosylation	A: 0.826 B: 0.754	0.161 (NS)
Stage IV **Set A: WELCA Set** **Set B: Dual Gender Lung Cancer Set**	α2-6 Sialylation	A: 0.887 B: 0.791	0.009
	β1-4 Branching	A: 0.917 B: 0.802	0.002
	β1-6 Branching	A: 0.907 B: 0.810	0.008
	Outer-Arm Fucosylation	A: 0.822 B: 0.777	0.307 (NS)

[a]The WELCA set was compared to the Dual Gender Lung Cancer Set and Stage I Only Lung Cancer Set (also dual gender). *N*-values of each group are shown in **Table 2.2**. Actual ROC curves are shown in **Fig. 2.4**.

[b]"NS" indicates no significant difference between the two compared ROC curves. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.013$ (NS), $p < 0.013$ (*), $p < 0.003$ (**), $p < 0.0003$ (***).

Table 2.6. Comparison of Top Performing Glycan Nodes in Male vs. Female Patients with Early Stage Lung Cancer.

Lung Cancer Sets[a]	Glycan Feature	A: ROC AUC of set A B: ROC AUC of set B	*p*-value of Delong's test for two ROC curves[b]
Dual Gender Lung Cancer Set Stage I Male vs Female Set A: Male Patients (n = 14) vs Male Controls (n = 123) Set B: Female Patients (n = 6) vs Female Controls (n = 76)	2-linked Mannose	A: 0.618 B: 0.544	0.537 (NS)
	α2-6 Sialylation	A: 0.646 B: 0.553	0.393 (NS) [c]
	β1-4 Branching	A: 0.629 B: 0.539	0.464 (NS) [c]
	β1-6 Branching	A: 0.598 B: 0.579	0.870 (NS) [c]
	4-linked GlcNAc	A: 0.592 B: 0.537	0.686 (NS)
	Outer-Arm Fucosylation	A: 0.552 B: 0.702	0.206 (NS)
Dual Gender Lung Cancer Set Stage II Male vs Female Set A: Male Patients (n = 12) vs Male Controls (n = 123) Set B: Female Patients (n = 8) vs Female Controls (n = 76)	2-linked Mannose	A: 0.628 B: 0.579	0.719 (NS)
	α2-6 Sialylation	A: 0.626 B: 0.602	0.873 (NS)
	β1-4 Branching	A: 0.633 B: 0.618	0.918 (NS)
	β1-6 Branching	A: 0.594 B: 0.559	0.826 (NS)
	4-linked GlcNAc	A: 0.648 B: 0.564	0.533 (NS)
	Outer-Arm Fucosylation	A: 0.737 B: 0.615	0.382 (NS)
Stage I Only Lung Cancer Set Male vs Female Set A: Male Patients (n = 33) vs Male Controls (n = 28) Set B: Female Patients (n = 74) vs Female Controls (n = 45)	2-linked Mannose	A: 0.655 B: 0.547	0.247 (NS)
	α2-6 Sialylation	A: 0.585 B: 0.557	0.766 (NS)
	β1-4 Branching	A: 0.448 B: 0.563	0.219 (NS) [c]
	β1-6 Branching	A: 0.450 B: 0.616	0.0825 (NS) [c]
	4-linked GlcNAc	A: 0.650 B: 0.589	0.508 (NS)
	Outer-Arm Fucosylation	A: 0.632 B: 0.610	0.811 (NS)

[a]Comparisons are made for stage I and II of the Dual Gender Lung Cancer Set, and the Stage I Only Lung Cancer Set. Unpaired Delong's test or Bootstrap test are applied to compare two ROC curves.

[b]"NS" indicates no significant difference between the two compared ROC curves. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.0083$ (NS), $p < 0.0083$ (*), $p < 0.0017$ (**), $p < 0.00017$ (***).

[c]*p*-value is from Bootstrap test instead of Delong's test, because Delong's test should not be applied to ROC curves with different directions and the stratification of Bootstrap is especially useful if groups are not balanced.

Table 2.7. Correlation Between Age and the Top Performing Glycan Nodes in the WELCA Cases (All stages) and, Separately, Controls.

Case/Control[a]	Glycan Feature	Correlation coefficient (r)	*p*-value of Spearman's rank correlation[b]
Case n = 208	2-linked Mannose	0.102	0.168 (NS)
	α2-6 Sialylation	0.073	0.324 (NS)
	β1-4 Branching	0.072	0.329(NS)
	β1-6 Branching	0.091	0.217 (NS)
	4-linked GlcNAc	0.030	0.681 (NS)
	Outer-Arm Fucosylation	0.148	0.044 (NS)
Control n = 207	2-linked Mannose	0.039	0.577 (NS)
	α2-6 Sialylation	0.075	0.283 (NS)
	β1-4 Branching	0.047	0.501 (NS)
	β1-6 Branching	0.103	0.142(NS)
	4-linked GlcNAc	-0.101	0.148 (NS)
	Outer-Arm Fucosylation	0.205	0.0031

[a]Spearman's rank correlation coefficients and *p* values are provided for the six top performing glycan features in all cases (n = 208) and controls (n = 207).

[b]"NS" indicates no significant correlation between age and the corresponding glycan feature. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.0083$ (NS), $p < 0.0083$ (*), $p < 0.0017$.

Table 2.8. Comparison of the Top Performing Glycan Nodes in Different Histological Types.

Histological Types[a]	Glycan Feature	A: ROC AUC of set A B: ROC AUC of set B	*p*-value of Delong's test for two ROC curves[b]
Adenocarcinoma vs Squamous cell carcinoma **Set A: Adenocarcinoma vs Controls** **Set B: Squamous cell carcinoma vs Controls**	2-linked Mannose	A: 0.854 B: 0.926	0.071 (NS) [c]
	α2-6 Sialylation	A: 0.878 B: 0.939	0.130 (NS) [c]
	β1-4 Branching	A: 0.908 B: 0.960	0.114 (NS) [c]
	β1-6 Branching	A: 0.906 B: 0.939	0.539 (NS) [c]
	4-linked GlcNAc	A: 0.877 B: 0.899	0.702 (NS) [c]
	Outer-Arm Fucosylation	A: 0.815 B: 0.861	0.608 (NS) [c]
Adenocarcinoma vs Large cell carcinoma **Set A: Adenocarcinoma vs Controls** **Set B: Large cell carcinoma vs Controls**	2-linked Mannose	A: 0.854 B: 0.860	0.957 (NS) [c]
	α2-6 Sialylation	A: 0.878 B: 0.817	0.647 (NS) [c]
	β1-4 Branching	A: 0.908 B: 0.828	0.586 (NS) [c]
	β1-6 Branching	A: 0.906 B: 0.808	0.402 (NS) [c]
	4-linked GlcNAc	A: 0.877 B: 0.869	0.934 (NS) [c]
	Outer-Arm Fucosylation	A: 0.815 B: 0.757	0.706 (NS) [c]
Squamous cell carcinoma vs Large cell carcinoma **Set A: Squamous cell carcinoma vs Controls** **Set B: Large cell carcinoma vs Controls**	2-linked Mannose	A: 0.926 B: 0.860	0.556 (NS)
	α2-6 Sialylation	A: 0.939 B: 0.817	0.406 (NS)
	β1-4 Branching	A: 0.960 B: 0.828	0.426 (NS)
	β1-6 Branching	A: 0.939 B: 0.808	0.332 (NS)
	4-linked GlcNAc	A: 0.899 B: 0.869	0.804 (NS)
	Outer-Arm Fucosylation	A: 0.861 B: 0.757	0.578 (NS)

[a]Comparisons are made for stage IV patients with various histological types of non-small cell lung cancer (NSCLC) vs. all controls. The n-values for the different histological sets are as following. Adenocarcinoma set: n = 70; Squamous cell carcinoma set: n = 8; Large cell carcinoma set: n = 5; Controls: n = 207. Unpaired Delong's test or Bootstrap test are used to compare two ROC curves.

[b]"NS" indicates no significant difference between the two compared ROC curves. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.0083$ (NS).

[c]*p*-value is from Bootstrap test instead of Delong's test, because the stratification of Bootstrap is especially useful if groups are not balanced.

Table 2.9. Stage-by-stage Comparison of Total Glycosylation with Individual Glycan Feature.

Stages[a]	Glycan Feature		ROC AUC	*p*-value of Delong's test for two ROC curves[b]
Stage I	A	β1-6 Branching	0.797	A vs B: 0.280 (NS) A vs C: 0.196 (NS) A vs D: 0.289 (NS)
	B	Total Hexoses	0.750	
	C	Total HexNAcs	0.730	
	D	Total Hexoses and HexNAcs	0.755	
Stage II	A	β1-6 Branching	0.770	A vs B: 0.024 (NS) A vs C: 0.091 (NS) A vs D: 0.017
	B	Total Hexoses	0.674	
	C	Total HexNAcs	0.627	
	D	Total Hexoses and HexNAcs	0.679	
Stage III	A	2-linked Mannose	0.843	A vs B: 0.938 (NS) A vs C: 0.676 (NS) A vs D: 0.196 (NS)
	B	Total Hexoses	0.844	
	C	Total HexNAcs	0.830	
	D	Total Hexoses and HexNAcs	0.860	
Stage IV	A	β1-4 Branching	0.917	A vs B: 0.021 (NS) A vs C: 0.159 (NS) A vs D: 0.280 (NS)
	B	Total Hexoses	0.892	
	C	Total HexNAcs	0.891	
	D	Total Hexoses and HexNAcs	0.907	

[a]For each stage, the individual top performing glycan node with the largest area under curve (AUC) value was selected to compare to total hexoses (sum of all hexose glycan nodes), total HexNAcs (sum of all HexNAc glycan nodes) and total Hexoses and HexNAcs (sum of all glycan nodes). A paired Delong's test was utilized to compare two ROC curves.

[b]"NS" indicates no significant difference between the two compared ROC curves. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.017$ (NS), $p < 0.017$ (*), $p < 0.0033$ (**).

Table 2.10. Survival Prediction by the Top Performing Glycan Nodes in All Stages, Stage III and IV Combined, Stage III Only and Stage IV Only.

Stage Involved	Glycan Feature	Cox proportional hazards regression model[a]			
		p-value[b]	Hazard Ratio	Lower bound at 95% CL	Upper bound at 95% CL
All stages n = 197	2-linked Mannose	0.0003	2.39	1.49	3.83
	α2-6 Sialylation	0.0002	2.48	1.53	4.03
	β1-4 Branching	< 0.0001	2.70	1.66	4.41
	β1-6 Branching	0.0002	2.54	1.57	4.12
	4-linked GlcNAc	0.0066	1.99	1.21	3.26
	Outer-Arm Fucosylation	< 0.0001	2.75	1.70	4.42
Stage III & IV n = 138	2-linked Mannose	0.0059	2.09	1.24	3.52
	α2-6 Sialylation	0.0029	2.16	1.30	3.58
	β1-4 Branching	0.0014	2.29	1.38	3.82
	β1-6 Branching	0.0014	2.29	1.38	3.82
	4-linked GlcNAc	0.0148 (NS)	1.98	1.14	3.42
	Outer-Arm Fucosylation	0.0011	2.45	1.43	4.18
Stage III n = 44	2-linked Mannose	0.3291 (NS)	1.69	0.59	4.81
	α2-6 Sialylation	0.1551 (NS)	2.12	0.75	5.99
	β1-4 Branching	0.5653 (NS)	1.35	0.49	3.75
	β1-6 Branching	0.1685 (NS)	2.04	0.74	5.65
	4-linked GlcNAc	0.2910 (NS)	1.68	0.64	4.42
	Outer-Arm Fucosylation	0.0769 (NS)	2.61	0.90	7.58
Stage IV n = 94	2-linked Mannose	0.0131 (NS)	2.19	1.18	4.05
	α2-6 Sialylation	0.0051	2.42	1.30	4.50
	β1-4 Branching	0.0011	2.82	1.51	5.26
	β1-6 Branching	0.0032	2.53	1.36	4.67
	4-linked GlcNAc	0.0220 (NS)	2.19	1.12	4.31
	Outer-Arm Fucosylation	0.0081	2.31	1.24	4.31

[a]Cox proportional hazards regression model *p* values and hazard ratios for the top quartile for each glycan node vs. all other quartiles combined, and lower and upper bound at 95% confident limits of hazard ratios are provided.

[b]"NS" indicates no statistically significance between hazard ratio and 1, representing no difference in the relative risk of death, comparing patients in the top quartile vs. all other quartiles of the respective glycan node. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: $p > 0.0083$ (NS), $p < 0.0083$ (*), $p < 0.0017$ (**), $p < 0.00017$ (***).

CHAPTER 3

A SPIN COLUMN-FREE APPROACH TO SODIUM HYDROXIDE-BASED GLYCAN PERMETHYLATION

3.1 Introduction

Glycans are complex carbohydrates that play important roles in numerous biological processes, such as intercellular trafficking, signal transduction, cell adhesion, endocytosis, immunosurveillance, and immune response initiation[160, 161]. Recently, much attention has been focused on the structural analysis and quantification of glycans in order to elucidate their biological roles in chronic human diseases, most notably cancer [10, 162, 163].

Permethylation is a useful derivatization method that confers several advantages with regard to the analysis of glycans—especially when combined with mass spectrometry (MS). It helps facilitate the determination of sequence and composition of monosaccharides in glycans, branching position and interglycosidic linkage information, and the presence of configurational and conformational isomers. Additionally, permethylation stabilizes sialic acid residues in glycan structures, improves measurement sensitivity, enhances separation of glycans by reversed-phase HPLC, and generates more predictable spectral patterns when subjected to tandem mass spectral (MS) analysis[73, 76]. Moreover, permethylation helps facilitate quantitative reproducibility when glycans are analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS[164].

The first carbohydrate permethylation methodology was reported by Purdie and Irvine in 1903[78]. Since that time, several improved permethylation procedures have

been developed. In 1964, Hakomori introduced a procedure with enhanced permethylation yield compared to previously described procedures by utilizing sodium methylsulfinyl carbanion (Na dimsyl) and iodomethane[80]. However, even with some modifications[81-83], the Hakomori procedure was still laborious and hazardous because the dimsyl reagent is extremely sensitive to moisture, air and carbon dioxide. In 1984, to circumvent these problems with the dimsyl reagent, Ciucanu and Kerek developed a procedure based on the reaction of glycans with iodomethane in dimethyl sulfoxide (DMSO) containing powdered sodium hydroxide (NaOH)[84]. This procedure has been successfully applied in various glycan structural studies and adopted or adapted by numerous laboratories [165]. The permethylation yield of this procedure for glycans containing N-acetylhexosamine (HexNAc) units, however, was not examined—perhaps because, as we have found, the yield of HexNAc residues is rather low with this procedure (unpublished data). More recently, a solid-phase spin column-based permethylation procedure was pioneered by Mechref, Novotny and coworkers[73, 85] for the analysis of small quantities of glycans, which could not be satisfactorily achieved by the Ciucanu procedure. In this procedure, the reaction of glycans with iodomethane is proposed to take place on the surface of NaOH beads that are packed in micro-spin columns. Since the solution-phase permethylated glycans are separated from the NaOH beads by centrifugation following the reaction, the issues associated with peeling reactions and oxidative degradation of glycans due to the presence of NaOH during the subsequent liquid/liquid extraction step are avoided. With its high derivatization efficiency and good analytical reproducibility, this procedure has demonstrated its utility for model glycoproteins and glycans derived from biomedical samples—even with low picomole to femtomole sample quantities. N-glycans, however, still need to be pre-isolated from glycoproteins and complex biological samples prior to

permethylation with the spin column procedure. In 2013, we adapted this permethylation method and incorporated trifluoroacetic acid (TFA) hydrolysis, reduction, and acetylation steps, to perform glycan linkage analysis directly from whole biofluids (such as serum and plasma) and homogenized tissue samples (**Fig. 3.1**)[74, 75]. With the additional steps, O-linked glycans, N-linked glycans, and glycolipids are all released and ultimately converted into partially methylated alditol acetates (PMAAs, **Fig. 3.2**) which are highly amenable to analysis by gas chromatography-mass spectrometry (GC-MS). Relative quantification of the resulting PMAAs (as surrogates for monosaccharide-and-linkage specific "glycan nodes") facilitates detection of changes in unique glycan features due to disease pathobiology[74]. Based on the unique branch points and linkages that they represent, some glycan nodes serve as 1:1 molecular surrogates of corresponding glycosyltransferases—the enzymes that facilitate construction of glycan polymers. Moreover, by analytically pooling together each chemically distinct glycan node from across all intact glycans in a complex sample, the approach condenses unique glycan structural features such as "core fucosylation", "6-sialylation", "bisecting GlcNAc", and "β1-6 branching" into single analytical signals (**Fig. 3.1**).

Herein we describe and document the performance characteristics of a new procedure that eliminates the need for a spin column when permethylating glycans from complex biological samples. Glycans containing HexNAc units and small glycan-containing samples from whole biofluids are readily permethylated by the new procedure. Permethylation efficiency is compared to standard spin column-based approaches based on both PMAA yield as well as intact permethylated O- and N-glycan yields. Finally, we evaluated the reproducibility of the spin column-free procedure

toward relative quantification of glycan nodes in blood plasma and assessed its clinical performance with regard to its ability to generate glycan nodes that distinguish breast cancer patient plasma from that of healthy, aged-matched controls.

3.2 Materials and Methods

3.2.1 Materials and samples

Materials: Galactopyranosyl-*β*-1,4-N-acetyl-D-glucosamine and 3-O-(2-Acetamido-2-deoxy-*β*-D-galactopyranosyl)-D-galactopyranose were obtained from Carbosynth (Compton, Berkshire, UK). Peptide-N-Glycosidase F (500 000 units/mL) was acquired from New England Biolabs (Ipswich, MA). Fetuin from fetal bovine serum and Bovine Ribonuclease B were purchased from Sigma-Aldrich (St. Louis, MO). Acetone was purchased from Avantor Performance Materials (Center Valley, PA). Methanol was obtained from Honeywell Burdick & Jackson (Muskegon, Ml). Acetonitrile and methylene chloride were acquired from Fisher Scientific (Fair Lawn, NJ). DMSO, iodomethane (99%, Cat. No. I8507), iodomethane-d_3 (99%, Cat. No. 176036), chloroform, trifluoroacetic acid (TFA), ammonium hydroxide, sodium borohydride, acetic anhydride, 2,5-Dihydroxybenzoic acid (DHB), sodium acetate and sodium hydroxide beads (20-40 mesh, Cat. No. 367176) were purchased from Sigma-Aldrich. Pierce spin colunns (900 μL volume) used in the 10 min permethylation procedure were obtained from ThermoFisher Scientific (Waltham, MA, Cat.No. 69705). Micro-spin columns (500 μL volume) for extended spin column permethylation procedure were purchased from Harvard Apparatus (Holliston, MA, Cat. No. 74-4420). GC-MS autosampler vials and Teflon-lined pierceable caps were obtained from ThermoFisher Scientific. GC consumables were acquired from Agilent (Santa Clara, CA); MS consumables were obtained from Waters (Milford, MA).

Plasma Samples: A set of 40 blood plasma samples, half from stage III-IV breast cancer patients and half from age-matched healthy women, were purchased from Bioreclamation IVT (Hicksville, NY). A 300-mL plasma sample from an individual donor was also obtained from Bioreclamation, which served as a quality control sample to ensure batch-to-batch quantitative reproducibility.

3.2.2 Experimental Procedures

3.2.2.1 Permethylation procedures

Spin column-free (SCF) permethylation procedure: Approximately 625 mg of sodium hydroxide (NaOH) beads were collected in a silanized 13 × 100 glass test tube, and then conditioned by one rinse with 350 μL of acetonitrile (ACN) followed by two rinses with 350 μL of DMSO. Thirty microliters (30 μL) of aqueous sample solution and 270 μL of DMSO were added to the preconditioned NaOH beads, followed by the addition of 105 μL iodomethane. The solution was then mixed thoroughly on a shaker at 650 rpm for 11 min. Next, to dissolve the NaOH beads and create a neutral to slightly acidic solution, 3.5 mL of acidified phosphate buffer was slowly added to each sample with continual mixing by manual swirling. To make the acidified buffer, concentrated hydrochloric acid (HCl) was mixed with 0.2 M phosphate buffer (composed of half 0.2 M Na_2HPO_4 and half 0.2 M NaH_2PO_4, pH 7) at a 1.3 : 2.2 volume ratio. Then 1.2 mL of chloroform was added and mixed thoroughly with the aqueous layer. After brief centrifugation (30s at 3000 g), the top layer (aqueous layer) of each sample was discarded, and replaced by a fresh 3.5-mL aliquot of 0.5 M NaCl solution in a 0.2 M sodium phosphate buffer (pH 7). Two additional liquid/liquid extraction cycles with 0.2 M sodium phosphate buffer (pH 7) were then performed. At last, the chloroform layer of each sample was collected and then dried under vacuum (Speed Vac) with no heat

applied (for further intact glycan analysis) or under a gentle stream of nitrogen (if the sample was to undergo glycan linkage analysis).

10 min spin column (SC 10 min) permethylation procedure: The following procedure was employed by Borges et al. in 2013[74] and 2016[75]. Sodium hydroxide beads were collected in a Pierce spin column (900 µL volume) up to 5 mm below the brim, and washed once with 350 µL of acetonitrile (ACN) followed by two rinses with 350 µL of DMSO. Nine microliters (9 µL) of aqueous glycan standard solution or a whole biofluid sample (e.g., blood plasma, serum, homogenized tissue, etc.) were mixed with 270 µL of DMSO in a 1.5 mL Eppendorf tube. The solution was added to the pre-conditioned NaOH beads in the plugged microfuge spin column, followed by the addition of 105 µL of iodomethane and immediate stirring. The sample mixture was then allowed to stand for 11 min with occasional gentle stirring. After reaction, the microfuge spin column was unplugged and spun for 30 s at 5000 rpm (1000 g in a fixed-angle rotor) to collect the sample solution, which was quickly transferred into 3.5 mL of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer (pH 7) within a silanized 13 × 100 glass test tube. NaOH beads were then washed twice by 300 µL of ACN, and the spin-throughs were all immediately transferred into the same silanized glass test tube. To each sample, 1.2 mL of chloroform was added to perform liquid/liquid extraction. With brief centrifugation (as above), the top layer (aqueous layer) was discarded and then replaced by a fresh aliquot of 3.5 mL of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer (pH 7). In total, three L/L extractions were performed for each sample, and finally the chloroform layer was extracted and then dried under vacuum (Speed Vac) with no heat applied (for further intact glycan analysis) or under a gentle stream of nitrogen (if the sample was to undergo glycan linkage analysis).

Extended spin column (SC 50 min) permethylation procedure: The following procedure was adapted from the protocol of Mechref et al.[166] Sodium hydroxide beads were added to a Harvard Apparatus micro-spin column (500 µL volume) up to 3 cm in height and washed twice with 50 µL of DMSO. In a 0.5 mL Eppendorf tube, 1.2 µL of aqueous glycan standard solution was mixed with 30 µL of DMSO. After mixing with 20 µL of iodomethane, the sample mixture was applied to the sodium hydroxide beads packed in a spin column. The sample mixture was allowed to stand for 30 min, followed by the addition of another 20 µL of iodomethane and 20 min incubation. The sample was collected by centrifugation at 200 g for 1 min after reaction, and the NaOH beads were washed twice by 50 µL of DMSO. The sample and two eluted volumes of DMSO were collected in a silanized 13 × 100 glass test tube. To perform L/L extractions, 3.5 mL of 0.5 M NaCl solution in 0.2 M sodium phosphate buffer (pH 7) and 1.2 mL chloroform were added to each sample. The chloroform layer was washed 3 times with the aqueous solution, then extracted and dried under vacuum (Speed Vac) with no heat applied (for further intact glycan analysis) or under a gentle stream of nitrogen (if the sample was to undergo glycan linkage analysis).

As carried out here, the final L/L extraction steps were necessary in order to facilitate subsequent analysis by GC-MS or MALDI-MS. As recently described elsewhere [167, 168], however, the effluent from the extended spin column procedure can simply be dried and reconstituted in LC-MS mobile phase provided that online desalting is carried out on the LC prior to elution of the permethylated glycans into the mass spectrometer. LC-MS was not employed in the studies described here, but it is unlikely that the product of the SCF procedure described above would be amenable to this L/L extraction-free, dry, dilute and shoot approach due to the high concentration of

salts that are present following dissolution of the NaOH beads with acidified phosphate buffer.

3.2.2.2 Glycan linkage analysis

Permethylation: Permethylation was carried out following one of the procedures listed above.

Hydrolysis, reduction, and acetylation: Hydrolysis, reduction and acetylation were carried out as we have previously described[74, 75]. Two molar TFA was prepared and added (325 µL) to each test tube and allowed to react for 2 h at 120 °C. Samples were dried under a gentle stream of nitrogen at 73 °C. Next, 10 mg/mL sodium borohydride in 1 M ammonium hydroxide was freshly prepared and added (475 µL) to each sample, which was allowed to stand for 1 h in room temperature. To remove excess borate, 100 µL of methanol (MeOH) was added and dried under nitrogen, followed by 125 µL of 9:1 (v/v) MeOH : acetic acid and drying under nitrogen. To achieve efficient acetylation, 18 µL of deionized water was added to each test tube, thoroughly mixed and sonicated for 2 min to dissolve any precipitates. Then each sample was mixed with 250 µL acetic anhydride and heated at 50 °C for 10 min, followed by adding 230 µL of concentrated TFA and heating again at 50 °C for 10 min. To clean up the sample mixture, 1.8 mL of dichloromethane and 2 mL of deionized water were added to perform liquid/liquid extraction twice, with the aqueous layer discarded for each cycle. The organic layer of each sample was transferred to a silanized autosampler vial, and then dried under a gentle stream of nitrogen at 40 °C.

Gas chromatography-mass spectrometry: An Agilent Model A7890 gas chromatograph (equipped with a CTC PAL autosampler) coupled to a Waters GCT (time-of-flight) mass spectrometer was employed for analysis of PMAAs by GC-MS.

To each sample vial, 64 μL of acetone was added to reconstitute the PMAAs. One microliter (1 μL) was then injected onto a hot (280 ℃), silanized glass liner containing a plug of silanized glass wool (Agilent Cat. No. 5183-4647) at split ratio of 20. Samples were separated via chromatography over a 30 m DB-5ms GC column, facilitated by helium as carrier gas with a 0.8 mL/min constant flow rate. The GC oven was held at an initial temperature of 165 °C for 0.5 min, ramped to 265 ℃ at a rate of 10 °C/min, followed by immediately ramping to 325 °C at a rate of 30 °C/min and holding at 325 °C for 3 min. Sample components eluted from GC column into the TOF mass spectrometer equipped with an electron ionization source (70 eV, 250 °C). MS spectra were collected in positive-ion mode. Spectra from individual TOF pulses over a m/z range of 40-800 were summed every 0.1s. The mass spectrometer was tuned and calibrated daily using perfluorotributylamine to ensure reproducible relative abundances of EI ions and mass accuracy within 10 ppm.

3.2.2.3 Intact glycan analysis

Enzymatic release and reduction of N-glycans: The following procedure, including the amount of glycoprotein employed, was adapted from Hu et al. for release and reduction of N-glycans[166]. A 10 mg/mL stock solution of N-glycoprotein (Bovine Ribonuclease B) was prepared. A 1 μL aliquot of the stock solution was mixed with 9 μL GlycoBuffer 2 (10×, 50mM sodium phosphate buffer, pH 7.5). Next, the mixture was heated in boiling water for 10 min, and then chilled on ice. Peptide-N-Glycosidase F (PNGase F) stock solution (500,000 units/mL) was diluted ten times with GlycoBuffer 2, and added (1 μL) to each sample. The glycoprotein samples were then incubated at 37 °C for 18 h, and dried under vacuum with assisted heat at approximately 55 ℃.

A 10 mg/mL solution of sodium borohydride in 1 M ammonium hydroxide solution was freshly prepared and added (10 µL) to each sample, which was then incubated in a 60 °C water bath for 1 h. Samples were mixed with 10 µL of 5% acetic acid, and dried under vacuum with assisted heat. Next, HPLC-grade methanol was added (100 µL) to each sample and removed by drying under vacuum. To eliminate residual borate in samples, the last two steps were repeated five times.

Permethylation: Permethylation was carried out following one of the procedures listed above.

MALDI-TOF mass spectrometry: The dried permethylated samples were reconstituted in 6 µL of ethanol. A 10 mg/mL 2,5-Dihydroxybenzoic acid (DHB) matrix solution was made in 1 mM sodium acetate solution. To directly compare samples derivatized with the SCF vs. SC 50 min procedures, 1 µL of sample permethylated by the SC 50 min procedure with iodomethane and 1 µL of sample permethylated by the SCF procedure with iodomethane-d_3 were mixed into 2 µL of DHB matrix solution, making a solution with a total volume of 4 µL. Next, 2 µL of the mixed solution was transferred onto a MALDI target spot, and then dried under vacuum for 10 min.

A Bruker Daltonics, Autoflex III L200 MALDI-TOF mass spectrometer was utilized for intact glycan analysis. The instrument was used in linear, positive ion, delayed-extraction mode, with 'ionsource 1' at 20 kV, 'ionsource 2' at 18.45 kV, lens at 7.7 kV, 50 ns delayed extraction, deflection signal suppression up to m/z 500, and 1 GS/s sample rate. To ensure good ion counting statistics, at least 2500 laser-shots were signal averaged for each mass spectrum. The MALDI-TOF mass spectrometer was externally calibrated with peptide calibration standard II, a mixture of 9 peptides

supplied by Bruker Daltonics (Cat. No. NC9349683).

3.2.3 Data analysis

GC-MS data: Masslynx 4.1 program and Quanlynx 4.1 software were utilized to process all GC-MS data for glycan linkage analysis. For routine relative quantification of each glycan "node" from glycan standards or blood plasma, the 2-4 most abundant and/or diagnostic fragment ions[74, 75] were summed using a 0.15 Da extracted ion chromatogram mass window for quantification. The summed extracted ion chromatogram (XIC) peak areas were automatically integrated and quantified, then exported to a spreadsheet for further calculation. All integration events were evaluated manually to ensure accuracy.

Permethylation efficiency was determined by the fraction of ion signal derived from PMAAs representing undermethylated form(s) of a specific glycan node divided by this same value plus the ion signal derived from the PMAA representing the fully permethylated form of that particular glycan node. (For example, for analysis of the Gal-4-GlcNAc standard, ion signal from the PMAA representing terminal galactose (t-Gal) represented the fully permethylated hexose while all other galactose-related ion signals such as those from PMAAs representing 2-Gal, 3-Gal, 4-Gal, etc. represented undermethylated forms of the t-Gal residue.) Because the total ion current chromatogram (TIC) contained a few chromatographic signals from sample preparation-related non-glycan contaminants that, in some cases, co-eluted with PMAAs of interest, extracted ion chromatograms had to be employed for this analysis. As such, we ensured that ions extracted for analysis of glycan undermethylation (see **Table 3.1 and 3.2**) covered greater than 90% of the ion current derived from the corresponding glycan node.

MALDI-TOF MS data: MALDI MS data for intact glycan analysis were processed by Bruker FlexAnalysis 3.0 software. Mass spectra were baseline subtracted using the TopHat algorithm, and MS peaks were integrated baseline-to-baseline. Integration events were inspected manually and, if necessary, adjusted to baseline as appropriate. Peak intensities were exported to a spreadsheet for further calculation. Calculated masses of the permethylated O-glycans from fetuin and N-glycans from ribonuclease B are provided in **Table 3.3 and 3.4**.

To directly compare the permethylation efficiency of the spin column-free procedure (SCF) and extended spin column procedure (SC 50 min), the intensity ratio of mass spectral peaks from the two procedures (SCF/SC 50 min) was calculated for each intact glycan. Every experiment was also run with the labeling reagents switched (i.e., iodomethane used for the SCF procedure and iodomethane-d_3 used for the SC 50 min procedure), then all results were pooled together to ensure a lack of labeling bias.

Statistical Analysis Software: Software employed for statistical analysis included Microsoft Excel for the two-tailed Student's t-test to compare the experimentally determined mean SCF/SC ratio for each glycoform with the theoretical value of 1. Excel was also used for the Benjamini-Hochberg false discovery rate adjustment of *p*-values for the Wilcoxon rank sum test. GraphPad Prism 5 was utilized for Kruskal-Wallis followed by Dunn's posthoc tests. And SAS 9.4 was employed for the Wilcoxon rank sum test.

3.3 Results and Discussion

3.3.1 Optimization of permethylation reagent volumes and order of addition for the spin column-free (SCF) permethylation procedure

To maximize permethylation efficiency of the SCF procedure, several parameters were assessed, including total volume of water in the original sample and order of reagent addition. Holding the volume of DMSO and iodomethane constant, the effect of sample water volume on permethylation yield was investigated at total water volumes of 10 μL, 20 μL, 30 μL, 40 μL and 80 μL. These volumes are given assuming that the original sample is aqueous in nature and constitutes a fraction of the stated volumes. N-acetylactosamine (LacNAc), composed of terminal galactose (t-Gal) linked to the 4-position of N-acetylglucosamine (4-GlcNAc), was chosen as the glycan standard for the assessment of permethylation efficiency, since the permethylation yields of both hexose and N-acethylhexosamine (HexNAc) monomer units can be evaluated from it simultaneously. For each sample, 5 μg of LacNAc was dissolved into the volumes of water listed above, and then processed by spin column-free permethylation followed by hydrolysis, reduction and acetylation steps.

As shown in **Fig. 3.3**, the highest permethylation efficiencies for both the hexose and HexNAc units, were found in samples with total original-sample water volumes of more than 30 μL. For samples with 30 μL total volume, less than 2% of hexose and less than 1% of HexNAc were undermethylated. But for samples with total original-sample water volumes under 30 μL, the permethylation efficiencies were not as good, evidenced by higher mean values and significantly higher undermethylation fractions (especially for 10 μL) based on Kruskal-Wallis followed by Dunn's posthoc tests. Based on previous work published by Ciucanu and coworkers[77], a small amount of water added to the permethylation system improves the solubility of glycans in organic solvents, such as DMSO, which is critical for the permethylation reaction. However, too much water added to the sample can lead to undesired effects on the permethylation

efficiency, due to the oxidative degradation of glycans in basic solution. In accordance with the previous work by Ciucanu et al.[77], a paste of sodium hydroxide was found on the walls of the test tubes immediately following the 11-minute incubation with NaOH beads for samples with a total volume above 40 μL, which could potentially result in inconsistent permethylation yields. Therefore, 30 μL was selected as the optimal original-sample water volume and was used in subsequent experiments.

The order in which water/DMSO-solubilized glycan samples were exposed to either iodomethane or the NaOH beads was also evaluated with regard to its impact on permethylated glycan yield. Generally, iodomethane was added to samples before the NaOH beads, even though sometimes a color change appeared suggesting oxidative degradation of glycans might be taking place[77]. To evaluate the possibility of completely avoiding oxidative degradation, samples were added to NaOH beads first, followed by addition of iodomethane. Five micrograms (5 μg) of LacNAc were permethylated by the SCF permethylation procedure, followed by hydrolysis, reduction and acetylation (n = 6 each). To compare the degrees of oxidative degradation for the two orders of reagent addition, the total ion signals (raw summed peak areas of XICs representing > 90% of the total ion PMAA ion signal) of t-Gal and 4-GlcNAc, as well as the ratios of 4-GlcNAc to t-Gal were tested by Wilcoxon rank sum test. No statistically significant differences were found for the degree of oxidative degradation of hexose and HexNAc units between the two orders of reagent addition (p = 0.7922 for t-Gal, and p = 0.6623 for 4-GlcNAc); neither was the HexNAc/hexose ratio significantly different. Even so, to minimize the risk of oxidative degradation it is preferable to add water/DMSO-solubilized samples to NaOH beads first followed by iodomethane.

3.3.2 Permethylation efficiency of the SCF procedure relative to spin column-based approaches

To validate the permethylation efficiency of our spin column-free procedure, we compared it with two other published and prevalent spin column-based permethylation procedures. One spin column permethylation procedure with 10 min reaction time (SC 10 min) was developed by Kang et al[73], Goetz et al[169] and further modified by Borges et al[75], the details of which can be found in their recently published work. A second spin column permethylation procedure with 50 min reaction time (SC 50 min) was summarized in recent work of Hu et al.[166] 2-Acetamido-2-deoxy-D-lactose (Gal-4GlcNc, a.k.a. LacNAc) and 3-O-(2-Acetamido-2-deoxy-β-D-galactopyranosyl)-D-galactopyranose (GalNAc-3Gal) were chosen as the model glycan standards for the investigation, both of which are disaccharides composed of hexose and HexNAc residues but with switched positions (reducing or non-reducing end) of the two classes of monosaccharides. We processed 5-µg samples of glycan standard with the three different permethylation procedures in parallel, then performed the hydrolysis, reduction and acetylation procedures together (i.e., with all samples in the same batch) to reduce variability in steps other than permethylation.

The undermethylated fraction of each monosaccharide was calculated (refer back to Method section) and results were evaluated by Kruskal-Wallis followed by Dunn's posthoc tests (**Fig. 3.4**). The data revealed that the spin column-free procedure has comparable or slightly better permethylation efficiency for hexoses and HexNAcs in comparison with the two spin column procedures (**Fig. 3.4**). According to our previous work[75], HexNAc residues are more susceptible to pH change in the aqueous solution before liquid/liquid extraction and thus have a higher risk of undergoing

oxidative degradation during permethylation than hexoses. But as shown here (**Fig. 3.4**), the undermethylation of HexNAc residues by the spin column-free procedure can reach as low as 1.4% (for 4-GlcNAc) and 0.4% (for t-GalNAc), demonstrating optimal control of permethylation efficiency by the spin column-free procedure. The peak areas under the summed extracted ion chromatograms (XICs) of hexose and HexNAc residues (refer to Method section and **Table 3.1 and 3.2**) were compared by evaluating their ratios, as shown in **Fig. 3.4e** and **4f**. Even with apparently lower relative yields of HexNAc residues for SCF procedure compared to the other two spin column procedures, the HexNAc/hexose ratios for the SCF procedure were sufficiently consistent and reproducible, which is critical during the routine analysis of biological samples.

3.3.3 Analysis of intact glycans following SCF permethylation

Generally, permethylation is utilized by most researchers to facilitate analysis of intact glycans. The spin column-free permethylation procedure can also be applied to intact O-glycan analysis, as well as intact N-glycan analysis if coupled with the release of glycans from N-glycoproteins. Therefore, we characterized the efficiency of SCF *vs*. spin column permethylation procedures for both intact O- and N-linked glycans.

Fetuins are heavily glycosylated proteins that carry both O- and N-linked glycans. Within the permethylation step, O-glycans are cleaved from the protein by *β*-elimination, while N-glycans remain attached[169]. Three samples of 145 µg of Fetuin were processed by the extended spin column permethylation procedure (SC 50 min), a commonly used spin column-based procedure[166, 170-174]. In parallel, 3 samples were processed with the SCF procedure but were permethylated with iodomethane-d_3. Samples treated by the two procedures were re-consitituted in 6 µL ethanol, mixed at a 1:1 ratio in a DHB matrix solution, and 2 µL was spotted onto a MALDI target. All

three samples derivatized by the SCF procedure with iodomethane-d_3 were mixed in all possible combinations with the three samples processed by the SC 50 min procedure that employed unlabeled iodomethane, leading to a total of 9 combinations (3 × 3). To eliminate the possibility of labeling reagent bias, the two procedures were carried out again on the same samples but with the labeling reagents switched.

The SCF procedure effectively permethylated all three major types of O-glycans from Fetuin-A (**Fig. 3.5**), with O-linked tetrasaccharides ($NeuAc_2$-Hex-HexNAc) and hexasaccharides ($NeuAc_2$-Hex_2-$HexNAc_2$) permethylated at nearly the same efficiency as the more commonly employed spin column-based procedure [166]. Highly averaged MALDI-MS peak intensities of identical glycoforms processed with the two procedures were compared directly and are shown as the SCF/SC ratio in **Fig. 3.6**. An SCF/SC ratio of 1 would indicate identical permethylation efficiency and yield for the two procedures. As such, we used a two-tailed Student's t-test to compare the experimentally determined mean SCF/SC ratio for each glycoform with the theoretical ratio of 1 (which would indicate an equal yield from the two procedures), using equation 1[175].

$$t = \frac{\bar{x}-\mu}{s/\sqrt{N}} \qquad \text{(Eq. 1)}$$

where t is the t-statistic, $\bar{x}$ is the sample mean, μ is the theoretical population mean of 1, s is the sample standard deviation, and N is the number of observations. For the case of tetra- and hexasaccharides, the SCF/SC ratios were nearly 1, despite the fact that the t-statistic for the tetra-saccharide was slightly higher than the critical value of t (t_{crit}) at the 99% confidence level (CL). Relative to the SC procedure, the SCF procedure was found to result in significantly lower yields of O-linked trisaccharides (NeuAc-Hex-HexNAc), mean SCF/SC ratio = 0.50) (**Fig. 3.6**). The mechanism underlying this

phenomenon is not known, but it is not likely caused by less efficient permethylation in the SCF procedure (cf. **Fig. 3.4**). It could, however, be related to relatively higher solubility of the permethylated trisaccharide in the aqueous layers of SCF samples, which have a lower pH during extraction—potentially resulting in decreased extraction efficiency. For larger O-linked glycans (tetra- and hexasaccharides), the comparison demonstrated that the SCF procedure produces nearly the same permethylated glycan yield as the widely employed spin column procedure (SC 50 min)[166].

To investigate the effectiveness of the SCF procedure on permethylation of N-glycans in comparison with the spin column procedure, 10 µg of bovine ribonuclease B (RNase B), a well-known N-glycoprotein with only one N-glycosylation site, which bears one of four major high-mannose N-glycans was employed[176] (**Fig. 3.5**). Since only O-glycans are released from proteins by β-elimination during the permethylation step[169], traditional enzymatic cleavage steps were adopted to release N-glycans from the glycoprotein (see Methods section). Using the comparison strategy described above for O-glycans, the SCF/SC ratios for the four major RNase B N-glycans were compared to the theoretical value of 1 by a two-tailed Student's t-test. Calculated t-statistics were larger than t_{crit} at the 99% CL, indicating that the SCF procedure produces a higher yield of permethylated high-mannose N-glycans compared to a widely used-spin column procedure (SC 50 min) (**Fig. 3.6**). Notably, while low-microgram quantities of RNase B were required to facilitate consistently high signal-to-noise (S/N) levels for this quantitative comparison, the SCF-based analysis was also run with 100 ng of RNAse B to demonstrate the sensitivity of the new approach. The S/N ratio for all four glycans was greater than 5 (results not shown).

3.3.4 Application of the SCF procedure to human blood plasma

As shown by Borges et al in 2013[74], spin-column based permethylation of whole biofluids and homogenized tissues followed by hydrolysis, reduction and acetylation can be used to generate composite glycan linkage analysis data from all major classes of glycans—including not only O-linked glycans, but N-linked and lipid-linked glycans as well—without the need for glycan-releasing enzymes. Relative quantification of each hexose is carried out by dividing its summed XIC peak area by the sum of all hexose XIC peak areas; likewise, relative quantification of each HexNAc is carried out by dividing its summed XIC peak area by the sum of all HexNAc XIC peak areas. These normalized abundances (NAs) are then used for further evaluation of biological differences between samples from donors with and without disease[75]. To determine the reproducibility of the SCF permethylation procedure as applied to blood plasma, 9-µL aliquots of an EDTA blood plasma sample from an individual donor were processed by SCF permethylation followed by hydrolysis, reduction, acetylation and analysis by GC-MS as we have previously described[74]. Six 9-µL aliquots were processed and analyzed three different times (i.e., 18 samples total in three separate batches) to evaluate intra- and interassay reproducibility. Shown in **Table 3.5** are the intra- and inter-assay reproducibility of all glycan nodes for which the mean normalized abundances contribute more than 1% of the total hexose or HexNAc signal. Average intra-assay reproducibility was 9.7% and inter-assay reproducibility was 17.7%. Of these 16 routinely detected glycan nodes in plasma samples, 15 had intra- and inter-assay reproducibility of less than 20% for SCF procedure. These metrics of analytical precision are comparable to those of the 10-min spin column permethylation procedure, as reported by Borges et al in 2016 [75]. The quantitatively inconsistent glycan node, 4-Glc, typically constitutes about 5% of the total hexose signal, but is a node that is primarily derived from glycolipids—which may behave differently than N- and O-

linked glycans under the sample processing conditions applied.

3.3.5 Application of the SCF permethylation procedure to glycan node analysis of blood plasma from breast cancer patients

The new SCF permethylation procedure was coupled to hydrolysis, reduction and acetylation and applied to a cohort of 40 blood plasma samples from breast cancer patients (stages III-IV, n = 20) and age-matched female controls (n = 20). Samples were randomized and analyzed blindly in seven batches as we have previously described[75]. Three glycan nodes with between-batch reproducibility of 16 %CV or better (**Table 3.5**) were found to be significantly different in the late stage-breast cancer samples compared to controls, according to a Wilcoxon rank sum test and Benjamini-Hochberg false discovery rate adjustment procedure with *p*-values less than 0.05 (shown in **Fig. 3.7**). The receiver operating characteristic (ROC) c-statistics (AUCs) of the three glycan nodes were in the range of 0.75-0.85 (**Fig. 3.7**), demonstrating that 2,4-linked mannose (2,4-Man), 2,6-linked mannose (2,6-Man) and 4,6-linked GlcNAc (4,6-GlcNAc) may potentially be useful as markers for stage III-IV breast cancer. Interestingly, these three glycan nodes represent the sum total of β1-4 branching, β1-6 branching, and core fucosylation, respectively—as pooled together from all intact glycans within each sample.

3.4 Conclusions

A new spin column-free (SCF) permethylation procedure that reduces consumable costs and the number of experimental steps required for glycan permethylation was developed and applied to mass spectrometry-based intact-glycan analysis and linkage-analysis of glycans from whole biospecimens. The permethylation efficiency of the new SCF procedure reached as high as 98% for hexose residues and

99% for HexNAc residues, which are values comparable to or slightly higher than those of some widely-used spin column-based procedures. With regard to intact glycan analysis, the SCF procedure provided almost identical yields for permethylated O-glycans composed of more than three monosaccharides, and higher yields for the permethylated N-glycans examined. The SCF permethylation approach, combined with hydrolysis, reduction and acetylation facilitated glycan linkage analysis of pooled glycans from unfractionated blood plasma with good quantitative reproducibility. Furthermore, glycan linkage analysis with the SCF procedure identified specific, previously unreported glycan "nodes" representative of β1-4 branching, β1-6 branching, and core fucosylation as potential markers of stage III-IV breast cancer.

Figures

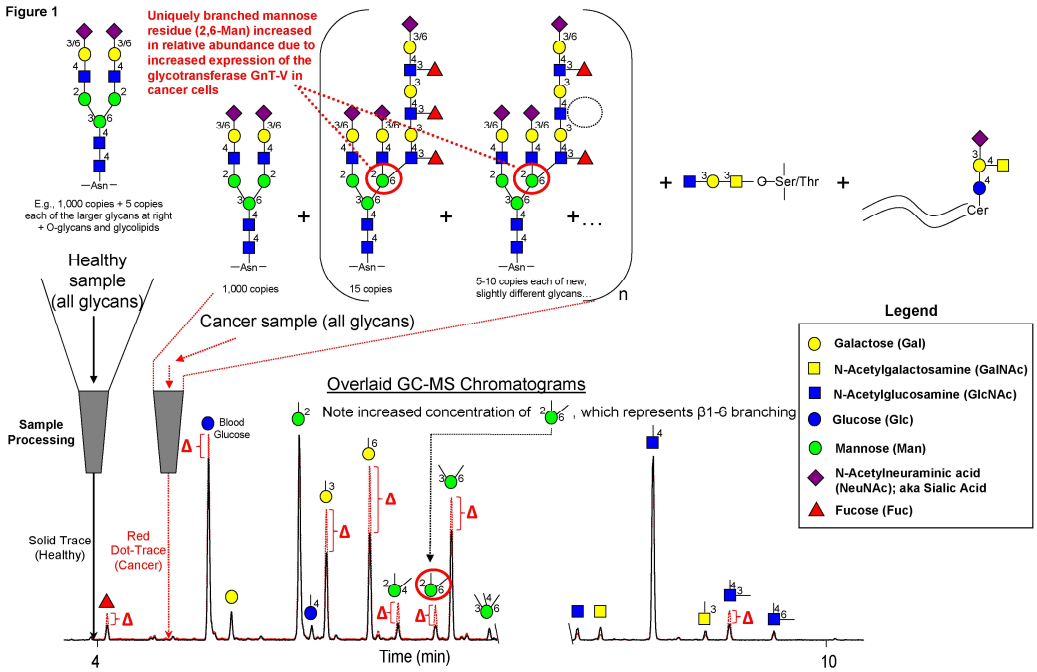


Figure 3.1: Conceptual overview of glycan linkage analysis from whole biospecimens. Intact normal and abnormal glycans including O-glycans, N-glycans and glycolipids, are processed and transformed into partially methylated alditol acetates (PMAAs, **Fig. 3.2**), each of which corresponds to a particular 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", "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.

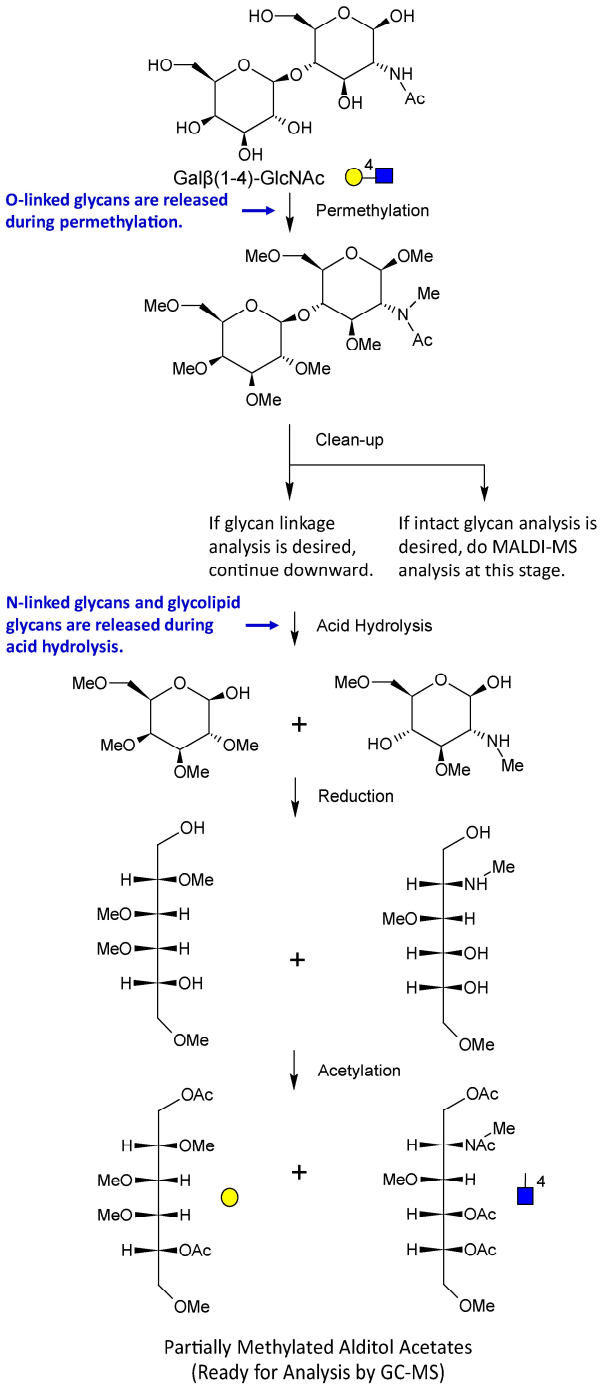


Figure 3.2: Molecular overview of the glycan linkage analysis procedure. The Gal-4-GlcNAc disaccharide is illustrated. For glycans from blood plasma and other biofluids, O-linked glycans are released during permethylation, while N-linked glycans and glycolipids are released during acid hydrolysis. The unique pattern of methylation and acetylation in the final partially methylated alditol acetates (PMAAs) corresponds to the unique "glycan node" in the original intact polymer and provides the molecular basis for separation and quantification by GC-MS.

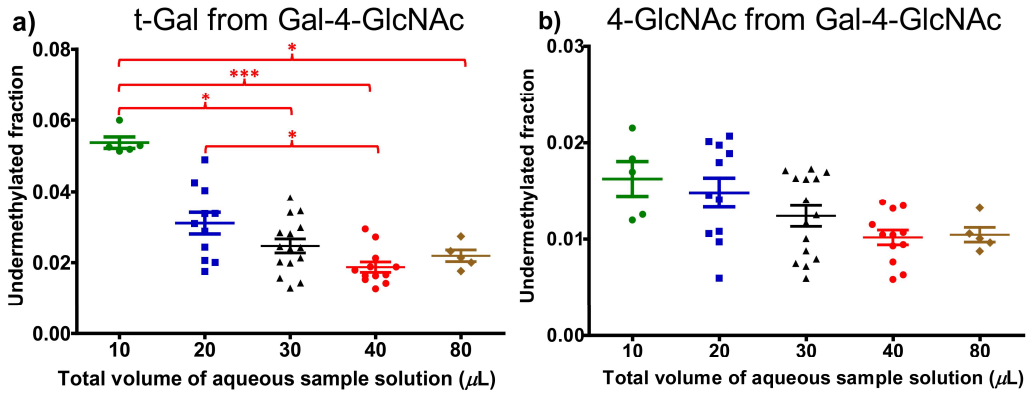


Figure 3.3: Effect of original-aqueous-sample volume on permethylation efficiency: (a) univariate distribution of the undermethylated fraction of t-Gal, (b) univariate distribution of the undermethylated fraction of 4-GlcNAc. N = 6 to 15 for each volume. Undermethylated fractions were determined as described in the Methods section. Kruskal-Wallis followed by Dunn's posthoc tests demonstrated significant differences in the degree of undermethylation at the tested volumes of water in the original samples. Error bars indicate mean ± standard error of the mean (SEM). Single asterisk (*) indicates significant differences between groups with $p \leq 0.05$. Three asterisks (***) indicate significant differences between groups with $p \leq 0.001$.

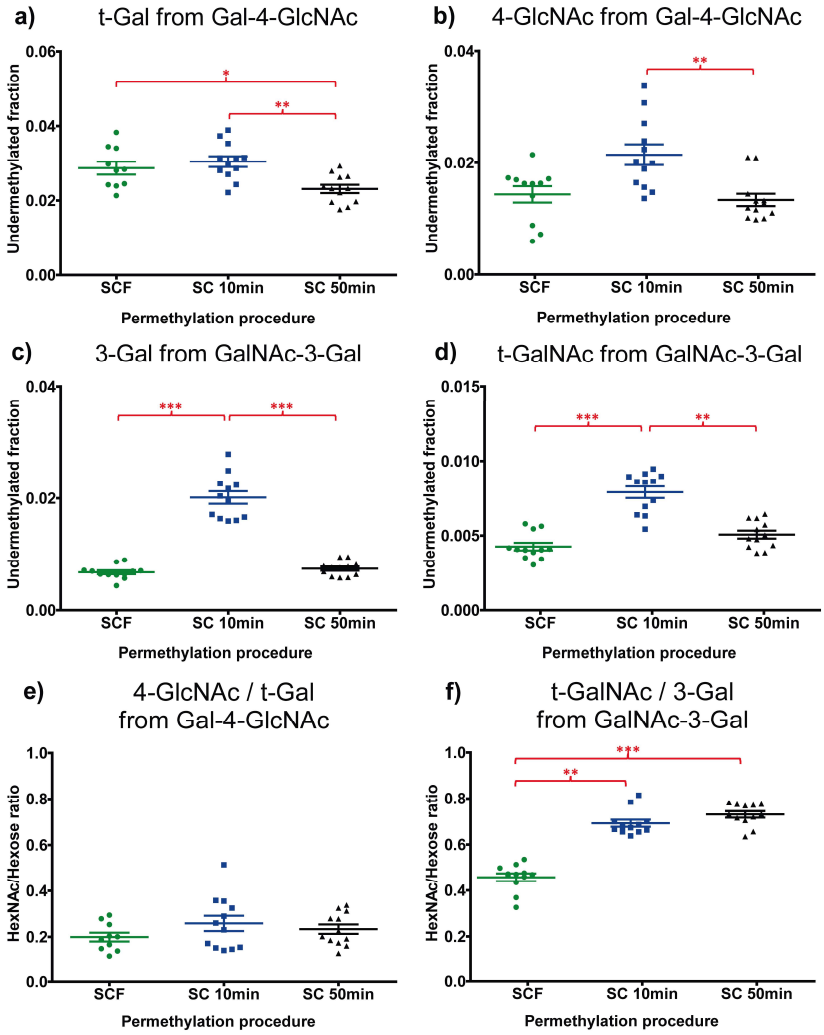


Figure 3.4: Evidence that the new spin column-free procedure has comparable permethylation efficiency for terminal/non-terminal hexose and terminal/non-terminal HexNAc relative to other prevalently used spin column procedures[74, 75, 166]. Undermethylated fractions were calculated and shown for (a) t-Gal from Galactopyranosyl-b-1,4-N-acetyl-D-glucosamine (Gal-4-GlcNAc), (b) 4-GlcNAc from Gal-4-GlcNAc, (c) 3-Gal from 3-O-(2-Acetamido-2-deoxy-b-D-galactopyranosyl)-D-galactopyranose (GalNAc-3Gal), and (d) t-GalNAc from GalNAc-3Gal. Peak areas under summed extracted ion chromatograms (XICs) of the HexNAc species were compared with that of hexose species, for both glycan standards, and the ratios were shown in (e) 4-GlcNAc/t-Gal, and (f) t-GalNAc/3-Gal. N = 12 for each permethylation procedure. Kruskal-Wallis followed by Dunn's posthoc tests demonstrated significant differences in the permethylation procedures. The SCF procedure produces a slightly higher undermethylated fraction than the extended SC procedure (SC 50 min) for t-Gal (a), but similar or lower undermethylated fractions than the two SC procedures for 4-GlcNAc (b), 3-Gal (c) and t-GalNAc (d). Error bars indicate mean ± standard error of the mean (SEM). Single asterisk (*) indicates significant differences between groups with $p \leq 0.05$. Two asterisks (**) indicate significant differences between groups with $p \leq 0.01$. Three asterisks (***) indicate significant differences between groups with $p \leq 0.001$.

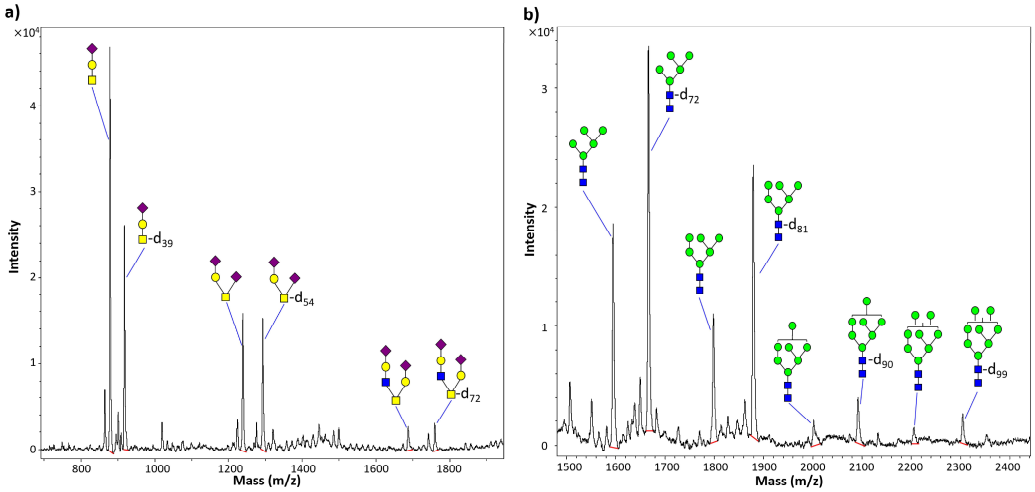


Figure 3.5: Permethylation of intact glycans by the spin column-free procedure (SCF) and extended spin column procedure (SC 50 min)[166]. a) O-glycans of fetuin from fetal bovine serum; b) N-glycans of bovine ribonuclease B. Glycan structure symbols indicate the most likely (isomerically ambiguous) corresponding glycan permethylated by the SC 50 min procedure with unlabelled iodomethane, or by the SCF procedure with iodomethane-d_3 (as indicated by "-d_n" labels where the subscripts represent the number of hydrogen atoms replaced by deuterium atoms in each glycan structure). Red lines under the peaks representing permethylated glycans indicate integration baselines. Calculated masses of the permethylated glycans depicted are provided in **Table 3.3 and 3.4**.

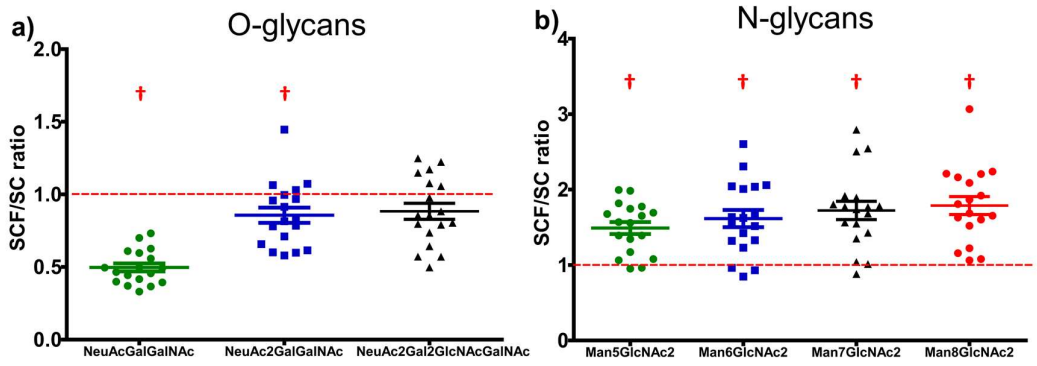


Figure 3.6: Ratio of intact, permethylated glycan yields produced by the SCF procedure relative to a commonly employed SC procedure (SC 50 min) [166]: (a) O-glycans from Fetuin and (b) N-glycans from bovine ribonuclease B. For each glycoform, the highly-averaged intensities of the two MALDI-MS peaks corresponding to the two permethylation procedures conducted with light and heavy labeling reagents were compared and presented here as the SCF/SC ratio. Two-tailed student's t-test comparing the SCF/SC ratios to the theoretical value of 1 demonstrated significant differences between the two procedures as indicated ($p < 0.05$). Error bars indicate mean ± standard error of the mean (SEM). † indicates significant differences between a group mean and the theoretical value of 1.

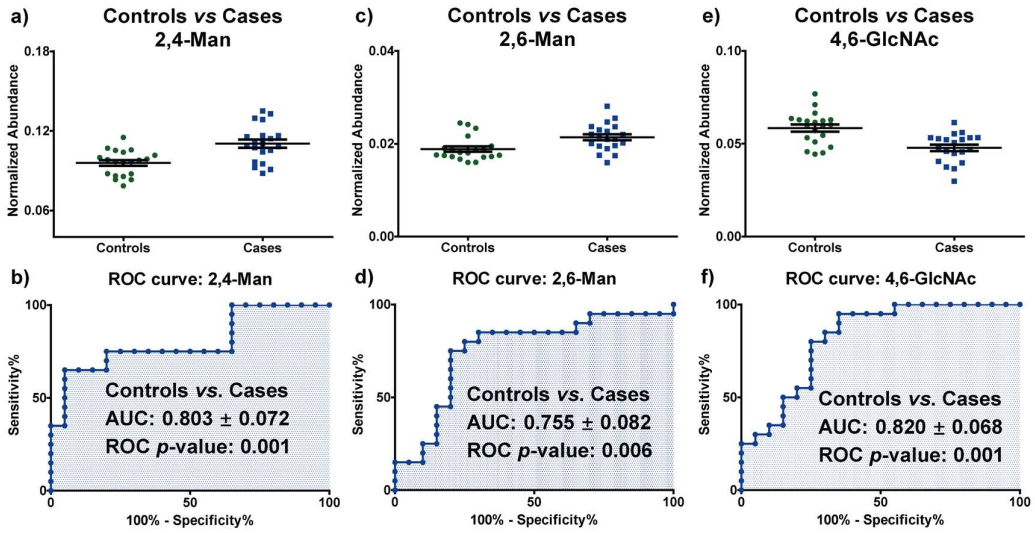


Figure 3.7: Application of the SCF procedure to glycan node analysis [74, 75] of clinical samples. Data shown were from blood plasma acquired from 20 stage III-IV breast cancer patients and 20 age-matched healthy women. Receiver operating characteristic (ROC) analysis of 20 glycan nodes indicated that three glycan nodes have the potential to distinguish breast cancer patients from controls: (a) univariate distribution of 2,4-Man, (b) ROC curve for 2,4-Man, (c) univariate distribution of 2,6-Man, (d) ROC curve for 2,6-Man, (e) univariate distribution of 4,6-GlcNAc, and (f) ROC curve for 4,6-GlcNAc. Data are based on normalized abundance (NA), which is defined as the summed XIC peak area of each glycan node divided by the sum of all other hexose (or HexNAc) summed XIC areas. For all three glycan nodes, $p < 0.01$ and AUC (area under curve) > 0.75 for ROC curves. Error bars indicate mean ± standard error of the mean (SEM).

Tables

Table 3.1. Extracted Ions for Disaccharide Gal-4-GlcNAc (± 0.15 Da).

Glycan Nodes	Extracted Ions (m/z)
t-Gal	59.05+71.05+87.04+101.06+117.05+129.05+145.08+161.07+173.08+205.10
4-Gal	59.05+71.08+87.04+101.06+113.06+117.05+129.05+131.07+161.07+173.07+233.11
2-Gal	59.05+71.05+87.04+101.06+113.05+117.05+129.05+145.08+161.08+189.07+205.10
3-Gal	59.05+74.04+87.06+101.06+117.05+129.05+143.08+161.08+173.07+201.08+233.11
6-Gal	59.05+71.08+87.04+99.04+101.06+117.05+129.05+159.06+161.08+173.07+189.07+233.11
3,4-Gal	59.05+74.04+87.04+117.05+129.05+143.08+161.08
2,3-Gal	86.07+129.05+143.08+161.08+201.08+261.10
3,6-Gal	98.07+101.06+117.05+129.05+189.07
4-GlcNAc	74.06+98.06+116.04+142.08+158.08+233.11
3,4-GlcNAc	74.06+98.06+116.07+142.08+158.08
4,6-GlcNAc	74.06+98.07+116.07+142.08+158.08

Table 3.2. Extracted Ions for Disaccharide GalNAc-3-Gal (± 0.15 Da)[a].

Glycan Nodes	Extracted Ions (m/z)
3-Gal f	59.05+74.04+87.05+101.06+117.06+161.09+171.07+189.09
3-Gal p	59.05+74.04+87.05+99.05+101.06+117.06+129.06+143.08+161.09+173.09+201.08+233.11
3,4-Gal	87.05+99.05+101.06+117.06+129.06+143.07+161.09
2,3-Gal	86.07+129.06+143.08+161.08+201.08+261.11
3,6-Gal	98.05+101.05+117.06+129.06+189.06
t-GalNAc	74.06+87.05+98.06+101.06+116.08+129.06+145.09+158.09+205.12
3-GalNAc	99.09+116.08+129.05+158.09+171.09+197.09
4-GalNAc	98.07+116.08+129.05+158.09+233.11
6-GalNAc	74.06+98.07+116.08+129.05+158.09+189.09+203.1
3,6-GalNAc	74.06+98.07+116.08+129.05+158.09+171.09+189.09

[a]Both furanose and pyranose form of 3-Gal (3-Gal *f* and 3-Gal *p*) were measured and summed to obtain a total signal for 3-Gal.

Table 3.3. **Masses (as $[M+Na]^+_{avg}$) of Intact O-glycans from Fetuin.**

Labeling Reagent	Trisaccharide	Tetrasaccharide	Hexsaccharide
Iodomethane	$C_{38}H_{68}N_2O_{19}$	$C_{54}H_{95}N_3O_{27}$	$C_{74}H_{130}N_4O_{37}$
	$[M+Na]^+$: 879.94	$[M+Na]^+$: 1241.33	$[M+Na]^+$: 1690.82
Iodomethane-d3	$C_{38}H_{29}D_{39}N_2O_{19}$	$C_{54}H_{41}D_{54}N_3O_{27}$	$C_{74}H_{58}D_{72}N_4O_{37}$
	$[M+Na]^+$: 919.18	$[M+Na]^+$: 1295.66	$[M+Na]^+$: 1763.26

Table 3.4. **Masses (as $[M+Na]^+_{avg}$) of Intact N-glycans from Ribonuclease B.**

Labeling Reagent	$Man_5GlcNAc_2$	$Man_6GlcNAc_2$	$Man_7GlcNAc_2$	$Man_8GlcNAc_2$
Iodomethane	$C_{70}H_{128}N_2O_{36}$	$C_{79}H_{144}N_2O_{41}$	$C_{88}H_{160}N_2O_{46}$	$C_{97}H_{176}N_2O_{51}$
	$[M+Na]^+$: 1596.75	$[M+Na]^+$: 1800.97	$[M+Na]^+$: 2005.19	$[M+Na]^+$: 2209.41
Iodomethane-d3	$C_{70}H_{56}D_{72}N_2O_{36}$	$C_{79}H_{63}D_{81}N_2O_{41}$	$C_{88}H_{70}D_{90}N_2O_{46}$	$C_{97}H_{77}D_{99}N_2O_{51}$
	$[M+Na]^+$: 1669.19	$[M+Na]^+$: 1882.47	$[M+Na]^+$: 2095.75	$[M+Na]^+$: 2309.02

Table 3.5. Intra- and Interassay Reproducibility of Permethylation Procedures as Applied to Glycan Linkage Analysis of Blood Plasma.

Glycan Nodes	Average Intra-Assay CV%[a]		Inter-Assay CV%[a]	
	SCF[b]	SC 10 min[c]	SCF[b]	SC 10 min[c]
t-Fuc	5.3	6.6	17.8	12.3
t-Gal	3.6	7.8	10.3	7.7
2-Man	3.6	2.8	7.6	5.3
4-Glc	45.2	13.5	54.7	15.6
3-Gal	13.4	20.2	18.9	26.8
6-Gal	3.8	2.7	10.8	9.8
2,4-Man	7.6	6.4	13.2	11.2
2,6-Man	5.2	5.8	8.9	6.5
3,6-Man	8.9	13.2	16.5	15.7
3,4,6-Man	4.6	7.4	18.7	19.0
t-GlcNAc	3.2	5.0	6.6	4.9
4-GlcNAc	2.0	2.1	6.5	2.9
3-GalNAc	10.1	7.0	16.5	8.7
3,4-GlcNAc	5.7	7.3	14.8	11.5
4,6-GlcNAc	6.1	3.3	15.8	6.1
3,6-GalNAc	9.9	12.2	18.2	15.4

[a]Values represent %CV of total hexose or total HexNAc-normalized individual glycan nodes. All glycan nodes contributing at least 1% of the total hexose or HexNAc signal are listed.

[b]Data were acquired by one analyst on 7 separate days by the new spin column-free procedure.

[c]Data were acquired by 3 separate analysts on 3 separate days by the standard 10-minute spin column procedure[74, 75]. n = 6 samples per batch.

CHAPTER 4

PREANALYTICAL HANDLING AND STORAGE VARIABILITY IN A MULTISITE-COLLECTION STUDY SIGNIFICANTLY IMPACTS ΔS-CYS-ALBUMIN, A BLOOD PLASMA/SERUM INTEGRITY MARKER

4.1 Introduction

In the past two decades, significant advances have been made in bringing the issue of research biospecimen integrity to light[177, 178]. Recent studies have revealed that the pre-analytical variables (PAVs) pervading the collection, processing and storage of biospecimens may potentially generate samples with unknown or unrecognized integrity issues, thus leading to inaccurate and unreliable research results [88, 96, 179, 180]. To address the issue of biospecimen integrity, efforts have been made toward the development of standard operating procedures (SOPs) for sample collection, processing and storage[177, 181]. However, the universal adoption of a single rigorous SOP, while in principle appealing as a solution to the development of potential biospecimen integrity problems, can, in practice, become impractical or impossible for laboratory personnel to implement on a rigorous basis for every sample or in situations where infrastructure issues arise such as unexpected power losses (see case study presented recently by Jeffs et al.[102]). This problem is amplified in studies with multiple collection sites involved—each of which must contend with different personnel and unique infrastructure constraints. Moreover, to date, many pre-clinical research studies have relied on pre-existing, archived biospecimens, some of which lack detailed paper trails documenting important PAVs. In 2014, a majority of the 455 extramural grants sponsored by the U.S. National Cancer Institute (NCI, part of the National Institutes of Health) utilized pre-existing samples (63%); of these 455 grants,

107 of the projects (24%) relied on pre-existing blood plasma or serum samples[92]. Recently, using our recently developed ΔS-Cys-Alb marker, we discovered a major integrity discrepancy in a set of serum samples collected under NCI sponsorship and slated for distribution to other cancer researchers[102]. It would be naïve to assume that this was a single, isolated incident amongst all archived P/S samples slated for use in future research studies (or that may have already been used in published research studies). As such, there is a present and urgent need for accurate quality control/quality assurance (QC/QA) markers or tests to retrospectively assess the integrity of archived biospecimens—even those for which a nominally pristine paper trail may exist.

Blood plasma/serum (P/S) samples, which can be routinely and quickly collected with minimal invasiveness, are among the most prevalently employed biospecimens in clinical research, particularly biomarker discovery research. In 2014, approximately 40% of NCI-sponsored grants involved the use of serum or plasma[92]. Several potential quality control markers for P/S integrity were outlined by Betsou et al[178], including transferrin receptor[182], potassium[183], ACTH and BNP[93], soluble CD40L[184], vitamin C[185, 186] and E[187], MMP-7[188], and etc. For nearly all of these candidate markers, the indication of disrupted sample integrity is grounded on the apparent loss of target analytes beyond their associated reference range.

As previously reported[102, 189, 190], human serum albumin contains a single free cysteine residue that can undergo spontaneous disulfide bond exchange with free cystine, forming S-cysteinylated albumin (S-Cys-Alb) to a limited degree in vivo and to a much greater extent when P/S samples are exposed to thawed conditions—i.e., temperatures > -30 °C [90, 91, 191, 192]. Charge deconvoluted ESI-mass spectra of albumin that illustrate the *ex vivo* formation of S-Cys-Alb are provided in **Fig. 4.1**. In the

presence of atmospheric oxygen and trace quantities of catalytic transition metals (primarily copper), nearly all of the cysteine equivalents in P/S (including those tied up as cystine) can be consumed by the native, reduced form of albumin producing S-Cys-Alb [102]. The number of cysteine equivalents in P/S is, on average, only about 20% of the number of albumin equivalents[126-128]. As such, the fraction of albumin in the S-cysteinylated form within nearly all human P/S samples (which, *in vivo*, starts at ~ 20-30%[102], will increase by ~ 20% to 40-50% if P/S samples are exposed to thawed conditions for a long enough period of time (~ 60 days at -20 °C, ~ 25 days at 4 °C, ~ 4-5 days at room temperature, and ~ 18 hrs at 37 °C[102]).

Based on these facts, Jeffs et al recently developed an assay known as "ΔS-Cys-Alb" as a QC/QA marker of P/S integrity[102]. This assay measures the S-Cys-Alb fraction change (ΔS-Cys-Alb) between the original P/S sample (presented with either a known or unknown storage history) and an aliquot of the same sample intentionally driven to its maximum *ex vivo* oxidation (S-cysteinylated) state. This inexpensive and rapid assay requires only 10 µL of a P/S sample and employs a very simple dilute-and-shoot LC-MS based analysis. Since the associated rate law model for the formation of S-Cys-Alb at room temperature was also established[102], back-calculation of the approximate time at which a given P/S samples must have been exposed to the equivalent of room temperature conditions was enabled—making it possible to place an "exposure time stamp" on every sample.

As mentioned above, we applied the ΔS-Cys-Alb assay to a case study of nominally pristine serum samples collected under NCI-sponsorship and discovered a previously undisclosed biospecimen integrity discrepancy[102]. To evaluate the potential utility of this assay to reveal variations in biospecimen handling and storage conditions, we have now applied it to 413 clinical plasma samples from the Women

Epidemiology Lung Cancer (WELCA) study, that were collected under a single SOP but from 12 different collection sites and then sent to a single repository for long-term storage. Here we provide insights into the effects of initial P/S handling conditions on ΔS-Cys-Alb and evaluate the meaning of these results with regard to their implications for the stability of important clinical analytes within P/S samples collected in this manner.

4.2 Materials and Methods

4.2.1 Materials

Trifluoroacetic acid (TFA, Cat. No. 299537) and formic acid (Cat. No. 06440) were obtained from Sigma-Aldrich (St. Louis, MO). Acetone (Cat. No. L-16923) and HPLC grade water (Cat. No. L16978) were acquired from ThermoFisher Scientific (Waltham, MA).

4.2.2 Samples

EDTA plasma samples from stage I-IV lung cancer patients and age-matched controls of Women Epidemiology Lung Cancer (WELCA) study were collected at 12 different collection centers in France under the supervision of Dr. Isabelle Stücker. This study was approved by Institutional Review Board of the French National Institute of Health and Medical Research and by the French data Protection Authority (IRB-Inserm, no. 3888 and CNIL no. C13-52). As part of the Women Epidemiology Lung Cancer Study, all-female lung cancer patients were recruited between September 2014 and November 2016 in collection sites 2 to 14, and age-matched all-female controls were recruited between June 2015 and December 2016 in collection site 18. All women living in Paris and the lle de France area, newly diagnosed with lung cancer were

considered as eligible cases. Age matched controls were randomly sampled from women without a history of cancer living in the same area. All Peripheral blood samples were drawn and processed following a written standardized protocol[150]. Briefly, after transported to the laboratory at 4 °C, blood samples collected in tubes containing EDTA additive were spun for 15 min at 3000 rpm and 4 °C in a standard centrifuge. In each of the 12 collection centers, the collected plasma samples were aliquoted and periodically (almost every three months) transported in dry ice to the central repository for final storage at -80 °C. No freeze-thaw cycles occurred prior to shipment to Arizona State University (Borges lab) for analysis. Received plasma samples were aliquoted on ice and kept frozen at -80 °C prior to any analysis.

Sample handling information including collection site, pre-centrifugation delay (from blood drawing to centrifugation), post-centrifugation delay (from centrifugation to storage), total pre-storage delay time (from blood drawing to storage), and initial storage temperature prior to shipment to the central repository were gathered and tabulated from the 413 corresponding sample collection information sheets.

A 300-mL plasma sample from an individual donor was obtained from BioIVT, which served as a quality control sample to ensure batch-to-batch quantitative reproducibility.

4.2.3 Experimental procedures

Sample preparation: Plasma samples, run in random order, were simply diluted 1000-fold before injection onto an LC-MS. Generally, a 10 μL aliquot of plasma was thawed at room temperature. Once completely in liquid status, 0.5 μL of plasma was added to 500 μL of 0.1 % (v/v) TFA and mixed thoroughly by vortexing for 20 s. Then the freshly diluted plasma solution was injected onto the LC-MS instrument

immediately. To intentionally drive the degree of albumin S-cysteinylation to its maximum value, the residual 9.5 µL plasma sample was incubated at 37 ℃ for 18 hrs in a 600-µL Eppendorf snap-cap test tube. Afterward, 0.5 µL of the sample (in fully oxidized state) was diluted 1000-fold and then loaded onto LC-MS, following the same steps described above.

LC-ESI-MS Analysis: The separation and relative quantification of intact albumin proteoforms were performed on an Agilent 1260 Infinity II UHPLC connected to an Agilent 6530 Electrospray Ionization Quadrupole Time-of-Flight (Q-TOF) LC-ESI-MS instruement. A 10 µL of sample was loaded at 200 µL/min in 80% water containing 0.1% formic acid (Solvent A) / 20% acetonitrile containing 0.1% formic acid (Solvent B) onto an Optimize Technologies protein captrap configured for bi-directional flow on a 6 – port diverter valve. The same solvent composition was held for 1 minute at 200 µL/min to rinse the protein captrap. Following were stepwise ramps of the flow composition to elute the trapped albumin at 200 µL/min: 3.0 – 3.1 min, ramped to 65/35 A/B and held until 4.5 min; 4.5 – 4.6 min, ramped to 55/45 A/B then held until 7.5 min; 7.5 – 7.6 min, ramped to 20/80 A/B and held until 8.6 min; at last, ramped back to 80/20 A/B by 8.7 min. This stepwise gradient results in near complete chromatographic separation of albumin from apolipoprotein A-I. A 4-minute post-run program was implemented at 200 µL/min of 80/20 A/B to reconstitute the system with the solvent composition of sample loading. The 6 – port valve was initially set to the loading position (1-2) and then switched to the inject position (1-6) at 3 minutes, directing the eluate from the captrap to the mass spectrometer. The mass spectrometer was set to run in positive ion, TOF-only mode and to collect spectra in *m/z* 100 to 3200 range. ESI settings for the Dual AJS ESI capillary microflow nebulizer ion source were

as follows: VCap 5700 V, Capillary 8.193 μA, Nozzle Voltage (Expt) 2000 V, Chamber 25.87 μA, Drying Gas nitrogen 7 L/min at 325 ℃, Nebulizer nitrogen 45 psig, Sheath Gas 11 L/min at 250 ℃. Data were acquired in profile mode at a rate of 1 spectra/s.

4.2.4 Data Processing

Across the chromatographic peak apex of albumin (retention time 5.5-5.9 min), approximately 0.3 minute of recorded spectra were averaged. Then deconvolution by the MaxEnt algorithm was carried out on the electrospray ionization charge-state envelope with Agilent MassHunter Qualitative Analysis vB.07.00 software. The MaxEnt algorithm parameters were set as follows: mass range 60000.00 – 72000.00 daltons, mass step 1.0000 daltons, using limited *m/z* range 1000 – 2500 *m/z*, subtracting baseline with baseline factor 0.7, adduct as proton, peak height filter with peak signal-to-noise ratio >= 10.0, maximum number of peaks limited (by height) to the largest 100, calculating average mass using top 90% of peak height, minimum consecutive charge states 10, and minimum protein fit score 8. In the deconvoluted spectra, the peak heights of mass spectral peaks of interest were exported to a spreadsheet for further statistical analysis.

The fractional abundance of S-cysteinylated albumin (oxidized form, S-Cys-Alb) was calculated by dividing the height of the peak representing S-Cys-Alb by the sum of the peak heights for native albumin and S-Cys-Alb. The change of S-Cys-Alb fraction (ΔS-Cys-Alb) for each plasma sample was determined by subtracting the S-Cys-Alb of the sample in its original state from the S-Cys-Alb of the same sample in its fully oxidized state.

Statistical analysis was performed in GraphPad Prism 7: For the four groups with different initial handling conditions, outliers were removed by log-transformation

and the ROUT method at Q = 1%. Outlier-removed data were then reverse-transformed by taking the anti-log of each value. To identify differences among groups, the Kruskal-Wallis test at 95% confidence level followed by the Benjamini-Hochberg false discovery correction procedure at a 5% false discovery rate was carried out. Correlation of ΔS-Cys-Alb with various delays were evaluated by spearman's rank correlation in GraphPad Prism 7. Differences between lung cancer patients and controls were assessed by means of the Mann-Whitney test at a 95% confidence level. GraphPad Prism 7 was also used to plot all the figures.

4.3 Results

For all 413 blood plasma samples obtained from the WELCA study, ΔS-Cys-Alb was measured in 33 batches. Samples were analyzed in random order. A quality control (QC) EDTA plasma sample was aliquoted and measured in each batch and indicated reasonable total interassay precision (%CV = 12.4%; at an S-Cys-Alb value of 20%, this corresponds to an S-Cys-Alb variability of 2.5%). 42

4.3.1 Effect of Initial Handling Conditions on ΔS-Cys-Alb

The effects of initial storage temperature (prior to shipment to the central repository) and total pre-storage delay on ΔS-Cys-Alb were evaluated first. Post analysis, blood plasma samples were divided into four groups based on their initial handling conditions: 1) samples stored at -80 ℃ within 3 hours from blood drawing; 2) samples stored at -80 ℃ but exposed to 4 ℃ or higher temperature for longer than 3 but less than 16 hours before storage at -80 ℃ , and 3) samples stored at -80 ℃ but exposed to 4 ℃ or higher temperature for longer than 16 hours before storage at -80 ℃ , and 4) samples stored at -20 ℃ for 42 up to 456 days. As might be expected based on the known stability of ΔS-Cys-Alb[102], pronounced, statistically significant differences

were observed in each group compared to one another (**Fig. 4.2**). Since ΔS-Cys-Alb measures the difference in the S-Cys-Alb fraction between the original status and fully oxidized status of a sample, a smaller ΔS-Cys-Alb represents more *ex vivo* oxidation had occurred in the sample prior to any analysis. The ΔS-Cys-Alb values of samples stored temporarily at -20 ℃ were found to be significantly lower than those stored in -80 ℃, indicating a definitively higher *ex vivo* oxidation level of albumin and compromised sample integrity. Moreover, for all the samples stored initially at -80 ℃, longer pre-storage delay at 4 ℃ also led to strikingly decreased ΔS-Cys-Alb value, implying possibly jeopardized sample integrity.

Next, because detailed information on pre- and post-centrifugation delay in hours as well as total pre-storage delay were available for most samples, the correlation of ΔS-Cys-Alb to each of these delay time categories was assessed. Spearman's rank correlation analysis demonstrated that ΔS-Cys-Alb was significantly correlated with pre-centrifugation delay (**Fig. 4.3a**) and total pre-storage delay (**Fig. 4.3c**), both with a moderate negative relationship (r = -0.522 and -0.524, respectively). When focusing on correlation to post-centrifugation delay, samples with pre-centrifugation delay longer than 3 hours were excluded to circumvent this potentially confounding factor. Thus, no statistically significant correlation was discovered between ΔS-Cys-Alb and post-centrifugation delay (**Fig. 4.3b**)—probably due to the relatively short delay time (< 2 hrs) after centrifugation for these samples. Together, these data revealed that delay time prior to samples being stored at less than -30 ℃, primarily before centrifugation, can adversely affect the ΔS-Cys-Alb values of plasma samples.

As reported in our previous work[102], the theoretical reference range of ΔS-Cys-Alb in P/S for 99% of the population has been predicted as 11 – 39% (yellow shaded

area in **Fig. 4.2**), which was calculated from the already known human population average plasma concentrations of albumin, cystine and cysteine[126-128]. In total, 96.7% of the ideally handled plasma samples from WELCA study, those stored in -80 ℃ within 3 hours from blood drawing, exhibited ΔS-Cys-Alb between 11 – 39%, indicating that this predicted theoretical range is almost certainly valid with regard to its application to lung cancer patient plasma samples.

4.3.2 ΔS-Cys-Alb at different collection sites

The distribution of ΔS-Cys-Alb observed at different collection sites is depicted in **Fig. 4.4**. As described above, plasma samples of lung cancer patients were collected at sites 2 to 14, and that of controls were obtained only at site 18. Among all 12 sample collection sites, sites 5 and 11 were obviously distinguished from the other sites, with markedly lower ΔS-Cys-Alb values (**Fig. 4.4a**). These findings were aligned with the fact that plasma samples collected from site 11 were all temporarily stored in -20 ℃ before being sent to the central repository (from 43 to 271 days), while in site 5 all samples except one (with ΔS-Cys-Alb value as 0.163) were stored temporarily at -20 ℃ (for 42 to 456 days). ΔS-Cys-Alb values from site 18 were also distinctive—exhibiting a lower and more widely spread range of ΔS-Cys-Alb values (**Fig. 4.4a**). At each collection site, collected and initially processed plasma samples were periodically (almost every three months) transported to the central repository biobank for final storage at -80 °C. Yet, since the plasma samples of the 208 lung cancer patients were collected from 11 dispersed sites, less than three batches each of lung cancer samples were acquired from sites 2 to 14. However, for site 18, a total of eight batches of plasma samples were gathered and sent to the central repository, two of which (received by the central repository in September and October 2016, respectively) demonstrated

statistically significant lower range of ΔS-Cys-Alb (**Fig. 4.4b**).

4.3.3 Equivalent time exposed to room temperature

Using the rate law-based mathematical model of S-Cys-Alb formation in P/S *ex vivo* established by our research group[102], the approximate time that an average P/S specimen has been exposed to room temperature (23 °C) can be back-calculated from a measured ΔS-Cys-Alb value. Referring to the table provided by Jeffs et al [102] where the rate law model-predicted P/S ΔS-Cys-Alb values were listed for 1 to 8 days exposed to 23 °C, the equivalent time of exposure to room temperature of each WELCA plasma sample was determined and depicted in **Fig. 4.5**. Samples were color-coded by collection site. Most samples from site 5, 11 and some from site 18 stood out from the rest, particularly in the range where the equivalent time exposed to room temperature was longer than 1 day. Taken along with the site-specific sample handling conditions described above, these data reveal that samples stored at -20 °C (for 42 to 456 days) exhibited the lowest ΔS-Cys-Alb values and therefore the greatest exposure time to the equivalence of 23 °C (as far as this particular marker is concerned).

4.3.4 Comparison of ΔS-Cys-Alb between lung cancer patients and controls

The oxidized form of P/S albumin has been proposed and employed as a marker for conditions related to oxidative stress[193-196], including chronic liver and kidney diseases and diabetes mellitus[109]. Given the data provided here and in our previous publications describing the *ex vivo* oxidation (S-cysteinylation) of albumin, it is clear that neither S-Cys-Alb or ΔS-Cys-Alb should be put forth as biomarkers of disease in the absence of extremely rigorous control and documentation of preanalytical handling and storage conditions. Nevertheless, in a study like this one, we would be remiss to not compare S-Cys-Alb and ΔS-Cys-Alb in the control samples to that of the cancer

patient samples (**Fig. 4.6**). Statistically significant differences were detected for both S-Cys-Alb and ΔS-Cys-Alb, but the distributions of the cancer patient population contained the entire control group population within itself and the mean ± SD values of the two patient groups were quite close (S-Cys-Alb: lung cancer cases, 0.296 ± 0.107; controls, 0.316 ± 0.072; ΔS-Cys-Alb: lung cancer cases, 0.156 ± 0.079; controls, 0.136 ± 0.056).

4.4 Discussion

The results presented above suggest that ΔS-Cys-Alb is quite sensitive to most initial biospecimen handling and storage conditions involving plasma sample exposure to thawed conditions (> -30 ℃). They also confirm that ΔS-Cys-Alb is non-linearly and inversely correlated with the time delay prior to storage—including and perhaps especially before centrifugation. A vast majority of properly handled samples (stored in -80 ℃ within 3 hours from blood drawing) fell within the previously reported theoretical population reference range (11 – 39%) of ΔS-Cys-Alb. Recent findings by Jeffs et al.[102] in fresh P/S collected from non-acute cardiovascular patients have also empirically confirmed this range, with the measured ΔS-Cys-Alb values in 95% of fresh cardiac patient plasma fall in the range of 14 – 30%. (Further validation in fully healthy donors is presently underway.)

The thawed state exposure times documented on paper and evidenced by the ΔS-Cys-Alb assay in this study were categorized into four initial handling/storage conditions (as illustrated in **Fig. 4.2a**). Other investigators have reported clinical analytes that were unstable under each of these four conditions, as listed in **Table 4.1**. A total of 6, 18, 64 and 54 commonly measured clinical analytes were found to be unstable under the four categories of handling/storage conditions, respectively. The

range of unstable clinical analytes included metabolites (vitamins, fatty acid, etc.), peptides and proteins (Lipoproteins, Prostate Specific Antigens, etc.), hormones (Insulin, Leptin, etc.), cell-free DNA, and inorganic cations/anions (Potassium, Bicarbonate, etc.) (**Table 4.1**).

When viewed in line with the findings documented in **Table 4.1**, the results from this study showed that temporary storage of plasma at -20 ℃ at satellite collection sites has a major impact on the integrity of samples treated in this manner. Although P/S samples appear to be frozen at -20 ℃, they are only actually frozen when the temperature is below -30 ℃[90, 91]. Moreover, temporary storage at 4 ℃ for more than 3 hours is generally ill-advised, because many more analytes start to lose stability with delay times longer than 3 hours (**Table 4.1**).

In light of the observations from this study, the ΔS-Cys-Alb assay may serve as a valuable tool to assess P/S integrity—especially for P/S samples for which handling and storage paper trails do not exist or may be of poor quality. Furthermore, pristine paper trails are sometimes insufficient to guarantee that sample integrity is not compromised by incidents. As reported by Jeffs et al., in a set of serum samples for a stage I lung cancer study, collected under NIH-sponsorship by seasoned investigators with a well-defined SOP, an explicit integrity discrepancy between cases and controls was detected by the ΔS-Cys-Alb assay [102]. After discovery, the integrity discrepancy was ultimately disclosed in detail by the clinical investigators who had provided the samples. In short, the -80 °C freezers in which the control samples had been stored had experienced a power outage for about 3-4 days during a natural disaster. As shown in **Fig. 4.4b**, substantial integrity discrepancies were observed between batches

of samples from the same site in this study (site 18). The explanation for this fact has not yet come to light—but it has, at least, been documented empirically.

ΔS-Cys-Alb is a novel marker for P/S integrity. Because it requires LC/MS, many laboratories will not be equipped to run it. However, the easy dilute-and-shoot nature of the ΔS-Cys-Alb assay along with its lack of requirement for a high-end LC-MS instrument means that nearly every MS facility and every major clinical reference lab is readily capable of running it. Moreover, investigators should bear in mind that to make a credible evaluation on sample sets, it is not always necessary to analyze every single sample; rather spot-checking a subset of each sample set that has been handled/stored together will, in many cases, suffice to empirically document P/S specimen integrity.

4.5 Conclusion

A simple dilute-and-shoot, intact-protein LC-MS assay known as ΔS-Cys-Alb that quantifies cumulative exposure of P/S samples to their thawed conditions (> -30 °C) was employed to evaluate the effect of various initial handling and storage conditions on P/S integrity in a clinical study involving multisite sample collection. ΔS-Cys-Alb values were dramatically lowered under conditions of prolonged pre-processing/pre-storage delay times at 4 °C and at an elevated temporary storage temperature (-20 °C) prior to shipment to the central repository where samples were permanently stored at -80 °C. In accord with a previously established rate law for the *ex vivo* formation of S-Cys-Alb in P/S, ΔS-Cys-Alb values were found to be non-linearly correlated with delay time prior to centrifugation as well as prior to storage. In consideration of the important clinical analytes that were unstable under the four categories of initial handling/storage conditions, ΔS-Cys-Alb will likely find utility as a tool to access P/S integrity.

Figures

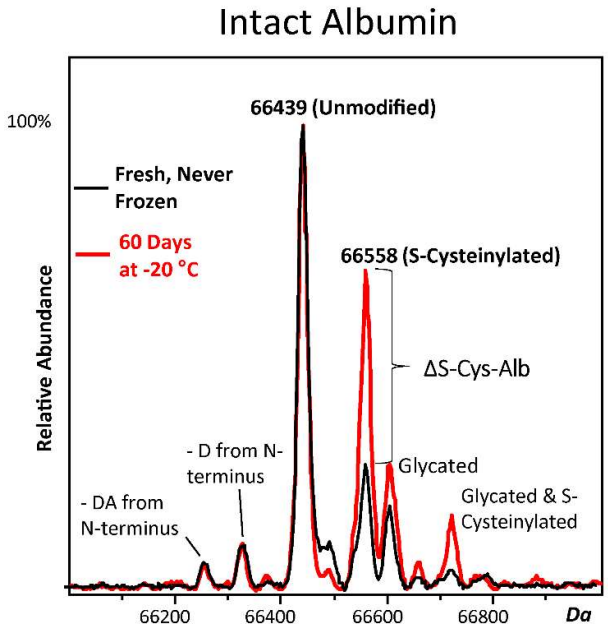


Figure 4.1. Charge deconvoluted ESI-mass spectra of albumin that illustrate ΔS-Cys-Alb. The black spectrum is from a fresh sample obtained immediately after centrifugation. The red spectrum is from the same sample stored for 60 days at -20 °C. As indicated, small fractions of albumin are N-terminally truncated, and both the native and S-cysteinylated forms may be glycated. This figure was originally published in Molecular & Cellular Proteomics[122]. © the American Society for Biochemistry and Molecular Biology.

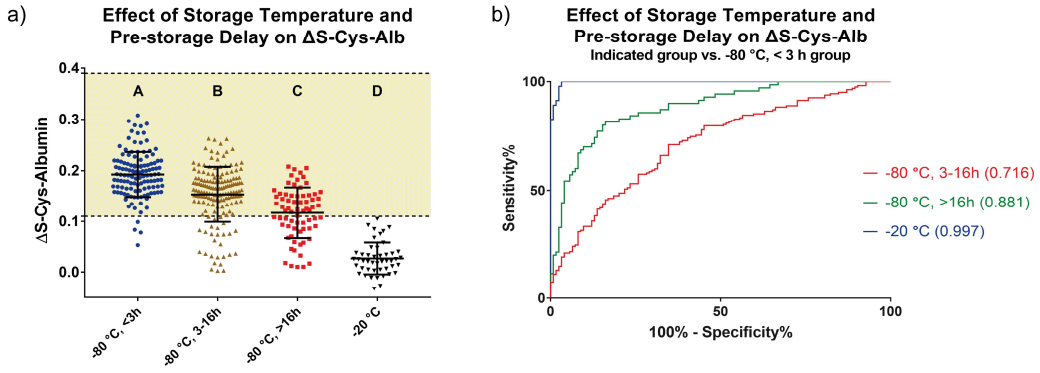


Figure 4.2. Effect of storage temperature and pre-storage delay on ΔS-Cys-Alb. (a) Univariate distribution of ΔS-Cys-Alb from four sample groups with different initial handling conditions. "-80 °C, <3h" represents the initial handling condition as samples were stored at -80 °C within 3 hours from blood drawing. Similarly, "-80 °C, 3-16h" and "-80 °C, >16h" stand for conditions under which samples were delayed before storage at -80 °C for 3 to 16 hours, and longer than 16 hours, respectively. Samples in the "-20 °C" group were initially stored at -20 °C for 42 up to 456 days prior to shipment to the central repository. Error bars represent mean ± SD, n = 124, 162, 70 and 46 for the four groups respectively. Eleven samples were excluded from the total 413 samples due to lack of initial handling information. The yellow shaded area indicates the theoretical reference range of ΔS-Cys-Alb in P/S (11 – 39%). The Kruskal−Wallis test was performed followed by the Benjamini−Hochberg false discovery correction procedure. The different letters above the data points in panel a) indicate statistically significant differences between groups ($p < 0.0001$ for each between-group comparison). (b) ROC curves for the three groups compared to the group with ideally handled samples (-80C, <3h). Areas under the ROC curves are provided in parenthesis next to the specified stages. For all three ROC curves, $p < 0.0001$.

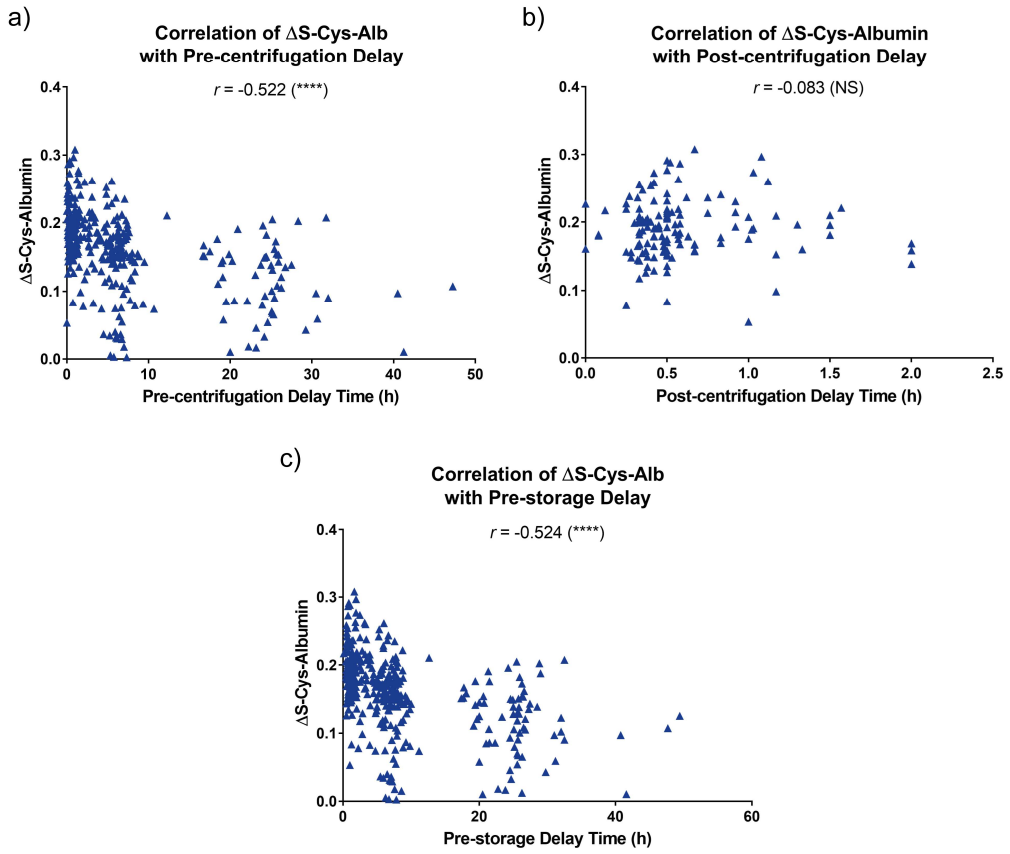


Figure 4.3. Correlation between pre- (a) and post-centrifugation delay (b), total pre-storage delay (c) and ΔS-Cys-Alb. Spearman's rank correlation coefficients are provided above the data points. "****" next to the coefficients in (a) and (c) indicates that the Spearman's rank correlation was statistically significant with $p < 0.0001$. "NS" next to the coefficient in panel b indicates a statistically insignificant Spearman's rank correlation.

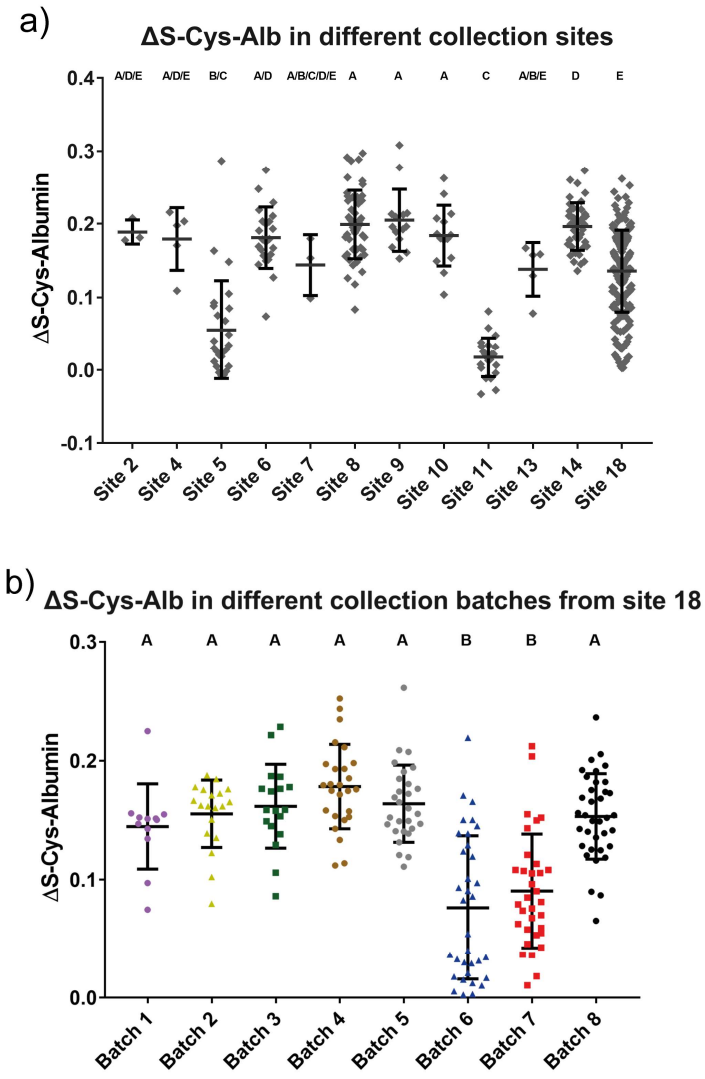


Figure 4.4. Univariate distribution of ΔS-Cys-Alb (a) at the different WELCA study collection sites, and (b) from different collection batches at site 18. The Kruskal−Wallis test was performed followed by the Benjamini−Hochberg false discovery correction procedure. The letters at the top of each set of data points within each panel demonstrate statistically significant differences between groups ($p < 0.05$); any overlap in letters indicates a lack of significant differences between groups ($p \geq 0.05$).

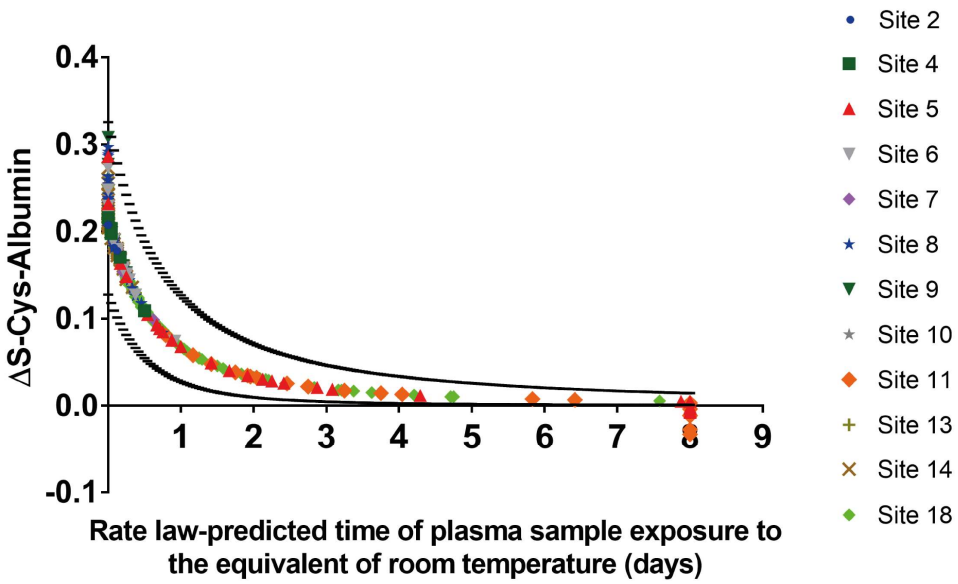


Figure 4.5. Estimate of plasma sample exposure times to the equivalent of 23 °C based on the rate law for formation of S-Cys-Alb[102]. Plasma samples collected at different sites were color-coded as illustrated in the legend on the right side. The curved line traced out by the data points represents the rate law-predicted ΔS-Cys-Alb value in a plasma sample with population-wide *average* initial concentrations of albumin, cystine, cysteine and copper [102]. The upper curved dashed line was calculated based on starting albumin and total copper concentrations of two SDs below the population mean but with a cystine concentration two SDs above the population mean; the lower curved dashed line was similarly calculated based on starting albumin and total copper concentrations of two SDs above the population mean but with a cystine concentration two SDs below the population mean. Thus, for each data point, the horizontal range between the two dashed black dashed lines represents the population-wide possible error for the calculated time of exposure to the equivalence of room temperature. Even though ΔS-Cys-Alb should, theoretically, always be positive, some slightly negative ΔS-Cys-Alb values were observed due to a modest degree of analytical error.

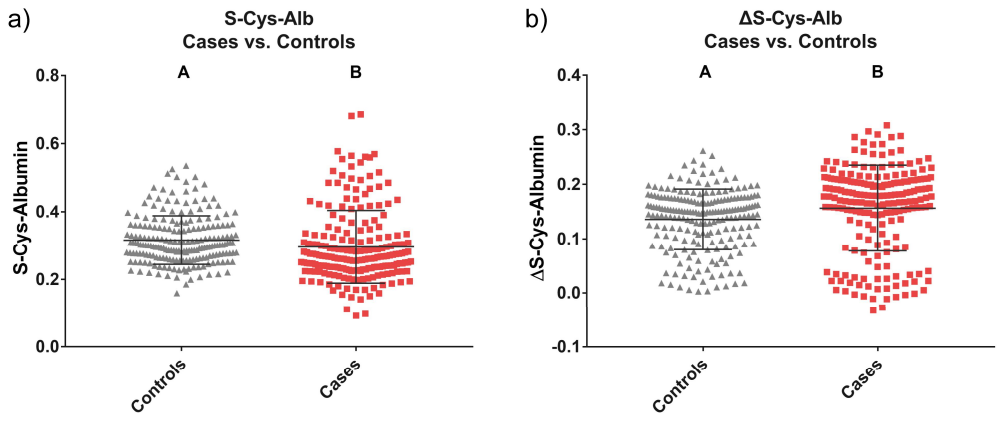


Figure 4.6. Univariate distribution of (a) S-Cys-Alb and (b) ΔS-Cys-Alb for lung cancer patients and controls. Mann-Whitney test was performed for both comparisons, with the letters above data points indicating statistically significant difference ($p < 0.0001$) between case and control groups. n = 207 for controls and n = 206 for lung cancer patients.

Tables

Table 4.1. Compilation of Studies Describing the Stabilities of Clinical Analytes in Blood Specimens under Various Initial Handling/Storage Conditions[a].

Initial handling/storage conditions	Biospecimen	Stable Analytes	Unstable Analytes	Reference
4 ° C for < 3 h	Serum		folic acid	197
	Plasma	ascorbic acid		185
	Serum		free prostate specific antigen (PSA)	198
	Serum		cell-free DNA	199
	Serum		ACTH	93
	Whole blood		potassium, glucose	200
4 ° C for 3-16 h	Serum	13 fatty acids		197
	Plasma	cell-free DNA		199
	Serum		free PSA	201
	Serum		transferrin receptor	182
	Plasma/Serum		ascorbic acid	185
	Whole blood		inorganic phosphorous, bicarbonate, lactate dehydrogenase, magnesium, activated partial thromboplastin time (APTT), C-telopeptide, adrenocorticotrophic hormone (ACTH)	200
	Plasma/Serum		potassium, bicarbonate, lactate, glucose, magnesium, APTT, osteocalcin, C-telopeptide	
4 ° C for > 16 h	Serum	C-reactive protein, retinol, ferritin	fatty acids (13)	197
	P/S	aldosterone, α-subunit total, α-subunit free, AVP, CRH, C-peptide, FSH, glucagon, GH, IGF-1, IGF-BP3, insulin, leptin, LH, PPP, PTH, prolactin, VIP		93
	Plasma		ACTH	
	Plasma	vitamin B12	folate	202
	Serum	total PSA		201
	Serum	total PSA, complex PSA		198
	Serum	apolipoprotein E		101
	Serum	cholesterol (IDL), triglycerides (VLDL, IDL), phospholipids (VLDL, IDL, LDL), proteins (VLDL, IDL, LDL, HDL)	VLDL cholesterol, LDL cholesterol, LDL triglycerides, HDL cholesterol, HDL triglycerides, HDL phospholipids	203
	Serum	palmitic acid, palmitoleic acid, stearic acid, oleic acid, linoleic acid, γ-linolenic acid, α-linolenic acid, dihomo-γ-linolenic acid, arachidonic acid, eicosapentaenoic acid, docosapentaenoic acid, docosahexaenoic acid, saturated fatty acid, mono	myristic acid	204

		unsaturated fatty acid, ω 3-polyunsaturated fatty acid, ω 6-polyunsaturated fatty acid		
	Plasma/Serum	C-telopeptides of type I collagen (CTX)		205
	Plasma		Cell-free DNA	199
	Plasma		catecholamines	206
	Whole blood		potassium	183
	Whole blood		lactate, mean corpuscular volume, mean corpuscular hemoglobin, follicle-stimulating hormone (FSH), estradiol, prolactin, folate, insulin, C-peptide, parathyroid hormone (PTH), osteocalcin	200
	Plasma/Serum		inorganic phosphorous, lactate dehydrogenase, FSH, estradiol, prolactin, folate, insulin, C-peptide, PTH	
	Plasma		sphingosine (d18:1), sphingadienine (d18:2), sphingosine (d16:1), glycerate, 5-hydroxyeicosatetraenoic acid (C20:trans[6]cis[8,11,14]4) (5-HETE), threonic acid, 3,4-dihydroxyphenylacetic acid (DOPAC), 3,4-dihydroxyphenylglycol (DOPEG)), Pentraxin-3, 10 peptides identified from fibrinogen, complement C4B and complement C3	207
-20˚C for 2 m to 5 y	Plasma	vitamin B_6, vitamin B_{12}, folic Acid, vitamin D	vitamin A, carotenoids, vitamin E	187
	Whole blood	vitamin B_1, vitamin, niacin		
	Serum	anti-dengue IgM, anti-dengue IgG		208
	Serum	C-reactive protein, creatine kinase, lactate dehydrogenase, AST, GGT, lipase, albumin, transferrin	amylase, alanine-amino transferase, alkaline phosphatase	209
	Serum	MMP-7	TIMP-1, VEGF, VEGF-receptor (Unstable after 3 m)	98
	Serum	total PSA	free PSA	201
	Serum	LDL cholesterol, LDL phospholipids, LDL proteins	ApoB, cholesterol (VLDL, IDL, HDL), triglycerides (VLDL, IDL, LDL, HDL), phospholipids (VLDL, IDL, HDL), proteins (VLDL, IDL, HDL)	203
	Serum	palmitic acid, palmitoleic acid, r-Linolenic acid, arachidonic acid, eicosapentaenoic acid, docosahexaenoic acid, saturated fatty acid, mono unsaturated fatty acid, ω 3-polyunsaturated fatty acid	myristic acid, stearic acid, oleic acid, linoleic acid, α-linolenic acid, dihomo-γ-Linolenic acid, Docosapentaenoic acid, ω 6-Polyunsaturated fatty acid	204
	Plasma/Serum	C-telopeptides of type I collagen (CTX)		205

	Serum		folate	210
	Plasma		catecholamines (Unstable after 1 m)	206
	Serum		free PSA	198
	Plasma/Serum		ascorbic acid (Unstable after 3 m)	185
	Serum		apolipoprotein E (Unstable after 3 m)	101
	Plasma		thromboplastin, prothrombin, thrombin, fibrinogen, coagulation factors II, V, VII, VIII, IX, X, XI and XII, protein C, protein S, antithrombin, plasminogen, von Willebrand factor, D-dimer (Unstable after 3 m)	99

[a]"h" stands for hour; "m" represents month; "y" represents year.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In this dissertation, the diagnostic and prognostic capacity of plasma glycan features in stage I-IV lung cancer—as represented by monosaccharide and linkage-specific glycan nodes—were validated in the WELCA case-control study. Significant elevation of α2-6 sialylation, β1-4 branching, β1-6 branching, outer-arm fucosylation and total N-glycosylation level, were observed in almost every stage of lung cancer relative to age-matched control groups. Alteration of glycan features in lung cancer was independent of smoking status, age and histological subtypes of lung cancer. Furthermore, the above six glycan features predicted all-cause mortality in lung cancer patients after adjusting for age, smoking status, and cancer stage. Early-stage detection was stronger than that previously observed by the Borges group, but this observation may have been related to the lack of smoking status-matching between cases and controls in the WELCA study. No gender-based differences were discovered in glycan features associated with lung cancer.

According to related studies previously reported by Borges group, some of the altered glycan features were also observed from a variety of other cancer types, such as pancreatic, ovarian, prostate, bladder and breast cancers, indicating that the alteration of many glycan features is arisen from inflammatory processes common to a majority of cancers. Further studies are needed to elucidate the mechanism of how the cancerous processes modulate changes in blood glycan biochemistry, as well as to shed light on the essential roles of these prevalent glycan features in the metastasis and progression of cancer.

To simplify the sample preparation steps and reduce consumable costs for

glycan permethylation, the first and most crucial step of glycan node analysis, a spin column-free (SCF) permethylation procedure was developed and applied to mass spectrometry-based glycan node analysis and intact-glycan analysis, with comparable or slightly better permethylation yields for N- and O-linked glycans as well as hexose and HexNAc residues. Furthermore, the new SCF procedure was approved to facilitate glycan node analysis of pooled glycans from unfractionated blood plasma with good quantitative reproducibility and dianostic capacity for late stage breast cancer. In the future, the performance of this new SCF procedure should be validated by applying it to cancer sets with larger sample size.

Although blood plasma glycans were discovered to be quite stable under a variety of sample storage conditions, the less-than-ideal conditions that expose P/S samples to their thawed conditions (> -30 °C) could have detrimental effects on a variety of clinical analytes. A simple dilute-and-shoot, intact-protein LC-MS assay known as ΔS-Cys-Alb was employed to evaluate the effect of various initial handling and storage conditions on P/S integrity in a clinical study involving multisite sample collection. It was revealed that prolonged pre-processing/pre-storage delay times at 4 °C and at an elevated temporary storage temperature (-20 °C) prior to the permanent storage at -80 °C could dramatically lower ΔS-Cys-Alb values. In line with a previously established rate law for the *ex vivo* formation of S-Cys-Alb in P/S, ΔS-Cys-Alb values were found to be non-linearly correlated with delay time prior to centrifugation as well as prior to storage. It is also pivotal to know that ΔS-Cys-Alb decays in a time frame that aligns well of plenty of other clinically important analytes present in P/S biospecimens. Therefore, ΔS-Cys-Alb will likely find utility as a robust, mechanistically understood tool to access P/S integrity and thus reduce the number of

misleading discoveries in biomarker research.

REFERENCES

[1] Kailemia, M. J., Ruhaak, L. R., Lebrilla, C. B., and Amster, I. J. (2013) Oligosaccharide analysis by mass spectrometry: a review of recent developments, *Analytical chemistry 86*, 196-212.

[2] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64-71.

[3] Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers, *Analytical chemistry 63*, 1193A-1203A.

[4] Zaia, J. (2010) Mass spectrometry and glycomics, *Omics: a journal of integrative biology 14*, 401-418.

[5] Zaia, J. (2004) Mass spectrometry of oligosaccharides, *Mass spectrometry reviews 23*, 161-227.

[6] Ruhaak, L., Zauner, G., Huhn, C., Bruggink, C., Deelder, A., and Wuhrer, M. (2010) Glycan labeling strategies and their use in identification and quantification, *Analytical and bioanalytical chemistry 397*, 3457-3481.

[7] Zauner, G., Deelder, A. M., and Wuhrer, M. (2011) Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics, *Electrophoresis 32*, 3456-3466.

[8] Higel, F., Demelbauer, U., Seidl, A., Friess, W., and Sörgel, F. (2013) Reversed-phase liquid-chromatographic mass spectrometric N-glycan analysis of biopharmaceuticals, *Analytical and bioanalytical chemistry 405*, 2481-2493.

[9] Harvey, S. R., MacPhee, C. E., and Barran, P. E. (2011) Ion mobility mass spectrometry for peptide analysis, *Methods 54*, 454-461.

[10] Varki, A., Kannagi, R., Toole, B., and Stanley, P. (2015-2017) Glycosylation Changes in Cancer, In *Essentials of Glycobiology [Internet]*. (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[11] Rini, J. M., and Esko, J. D. (2015-2017) Glycosyltransferases and Glycan-Processing Enzymes, In *Essentials of Glycobiology [Internet]*. (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[12] Taniguchi, N., Honke, K., and Fukuda, M. (2011) *Handbook of glycosyltransferases and related genes*, Springer Science & Business Media.

[13] Narimatsu, H. (2004) Construction of a human glycogene library and

comprehensive functional analysis, *Glycoconjugate journal 21*, 17-24.

[14] Grabenhorst, E., Nimtz, M., Costa, J., and Conradt, H. S. (1998) In Vivo Specificity of Human α1, 3/4-Fucosyltransferases III-VII in the Biosynthesis of LewisX and Sialyl LewisX Motifs on Complex-type N-Glycans COEXPRESSION STUDIES FROM BHK-21 CELLS TOGETHER WITH HUMAN β-TRACE PROTEIN, *Journal of Biological Chemistry 273*, 30985-30994.

[15] Ramakrishnan, B., Boeggeman, E., Ramasamy, V., and Qasba, P. K. (2004) Structure and catalytic cycle of β-1, 4-galactosyltransferase, *Current opinion in structural biology 14*, 593-600.

[16] Brodbeck, U., Denton, W., Tanahashi, N., and Ebner, K. E. (1967) The isolation and identification of the B protein of lactose synthetase as α-lactalbumin, *Journal of Biological Chemistry 242*, 1391-1397.

[17] Brew, K., Vanaman, T. C., and Hill, R. L. (1968) The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction, *Proceedings of the National Academy of Sciences of the United States of America 59*, 491.

[18] Hammond, C., Braakman, I., and Helenius, A. (1994) Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control, *Proceedings of the National Academy of Sciences 91*, 913-917.

[19] Hebert, D. N., Zhang, J.-X., Chen, W., Foellmer, B., and Helenius, A. (1997) The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin, *The Journal of cell biology 139*, 613-623.

[20] Ou, W.-J., Cameron, P. H., Thomas, D. Y., and Bergeron, J. J. (1993) Association of folding intermediates of glycoproteins with calnexin during protein maturation, *Nature 364*, 771.

[21] Degen, E., and Williams, D. B. (1991) Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility molecules, *The Journal of cell biology 112*, 1099-1115.

[22] Wang, X., Inoue, S., Gu, J., Miyoshi, E., Noda, K., Li, W., Mizuno-Horikawa, Y., Nakano, M., Asahi, M., and Takahashi, M. (2005) Dysregulation of TGF-β1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice, *Proceedings of the National Academy of Sciences 102*, 15791-15796.

[23] Maverakis, E., Kim, K., Shimoda, M., Gershwin, M. E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L. R., and Lebrilla, C. B. (2015) Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review, *Journal of autoimmunity 57*, 1-13.

[24] Bergstrom, K. S., and Xia, L. (2013) Mucin-type O-glycans and their roles in intestinal homeostasis, *Glycobiology 23*, 1026-1037.

[25] Fu, J., Wei, B., Wen, T., Johansson, M. E., Liu, X., Bradford, E., Thomsson, K. A., McGee, S., Mansour, L., and Tong, M. (2011) Loss of intestinal core 1–derived O-glycans causes spontaneous colitis in mice, *The Journal of clinical investigation 121*, 1657-1666.

[26] An, G., Wei, B., Xia, B., McDaniel, J. M., Ju, T., Cummings, R. D., Braun, J., and Xia, L. (2007) Increased susceptibility to colitis and colorectal tumors in mice lacking core 3–derived O-glycans, *Journal of Experimental Medicine 204*, 1417-1429.

[27] Fukuda, M., Hiraoka, N., and Yeh, J.-C. (1999) C-type lectins and sialyl Lewis X oligosaccharides: versatile roles in cell–cell interaction, *The Journal of cell biology 147*, 467-470.

[28] Munro, J. M., Lo, S. K., Corless, C., Robertson, M. J., Lee, N. C., Barnhill, R. L., Weinberg, D. S., and Bevilacqua, M. P. (1992) Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues, *The American journal of pathology 141*, 1397.

[29] Schnaar, R. L., and Kinoshita, T. (2015-2017) Glycosphingolipids, In *Essentials of Glycobiology [Internet]*. (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[30] van Eijk, M., Aten, J., Bijl, N., Ottenhoff, R., van Roomen, C. P., Dubbelhuis, P. F., Seeman, I., Ghauharali-van der Vlugt, K., Overkleeft, H. S., and Arbeeny, C. (2009) Reducing glycosphingolipid content in adipose tissue of obese mice restores insulin sensitivity, adipogenesis and reduces inflammation, *PLoS One 4*, e4723.

[31] Bremer, E. G., Schlessinger, J., and Hakomori, S.-I. (1986) Ganglioside-mediated modulation of cell growth. Specific effects of GM3 on tyrosine phosphorylation of the epidermal growth factor receptor, *Journal of Biological Chemistry 261*, 2434-2440.

[32] Uchida, Y., and Hamanaka, S. (2006) Stratum corneum ceramides: function, origins, and therapeutic applications, *Skin Barrier. New York: Taylor & Francis*, 43-64.

[33] Hoshi, T., Suzuki, A., Hayashi, S., Tohyama, K., Hayashi, A., Yamaguchi, Y., Takeuchi, K., and Baba, H. (2007) Nodal protrusions, increased Schmidt‐Lanterman incisures, and paranodal disorganization are characteristic features of sulfatide‐deficient peripheral nerves, *Glia 55*, 584-594.

[34] Lu, Y., Pelling, J. C., and Chaney, W. G. (1994) Tumor cell surface β1–6 branched oligosaccharides and lung metastasis, *Clinical & experimental metastasis 12*, 47-54.

[35] Granovsky, M., Fata, J., Pawling, J., Muller, W. J., Khokha, R., and Dennis, J. W. (2000) Suppression of tumor growth and metastasis in Mgat5-deficient mice, *Nature medicine 6*, 306.

[36] Seberger, P. J., and Chaney, W. G. (1999) Control of metastasis by Asn-linked, β1–6 branched oligosaccharides in mouse mammary cancer cells, *Glycobiology 9*, 235-

241.

[37] Lau, K. S., and Dennis, J. W. (2008) N-Glycans in cancer progression, *Glycobiology 18*, 750-760.

[38] Tanabe, K., Deguchi, A., Higashi, M., Usuki, H., Suzuki, Y., Uchimura, Y., Kuriyama, S., and Ikenaka, K. (2008) Outer arm fucosylation of N-glycans increases in sera of hepatocellular carcinoma patients, *Biochemical and biophysical research communications 374*, 219-225.

[39] Taylor-Papadimitriou, J., Burchell, J., Miles, D., and Dalziel, M. (1999) MUC1 and cancer, *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1455*, 301-313.

[40] Ogata, S., Maimonis, P. J., and Itzkowitz, S. H. (1992) Mucins Bearing the Cancer-associated Sialosyl-Tn Antigen Mediate Inhibition of Natural Killer Cell Cytotoxicity, *Cancer Research 52*, 4741-4746.

[41] Zhang, K., Sikut, R., and Hansson, G. C. (1997) A MUC1 Mucin Secreted from a Colon Carcinoma Cell Line Inhibits Target Cell Lysis by Natural Killer Cells, *Cellular Immunology 176*, 158-165.

[42] Wesseling, J., Van Der Valk, S. W., Vos, H. L., Sonnenberg, A., and Hilkens, J. (1995) Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components, *The Journal of cell biology 129*, 255-265.

[43] Wesseling, J., van der Valk, S. W., and Hilkens, J. (1996) A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by the membrane-associated mucin episialin/MUC1, *Molecular biology of the cell 7*, 565-577.

[44] Fu, C., Zhao, H., Wang, Y., Cai, H., Xiao, Y., Zeng, Y., and Chen, H. (2016) Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen, *HLA 88*, 275-286.

[45] Cheresh, D. A., Pierschbacher, M. D., Herzig, M. A., and Mujoo, K. (1986) Disialogangliosides GD2 and GD3 are involved in the attachment of human melanoma and neuroblastoma cells to extracellular matrix proteins, *The Journal of Cell Biology 102*, 688-696.

[46] Spiegel, S., and Fishman, P. H. (1987) Gangliosides as bimodal regulators of cell growth, *Proceedings of the National Academy of Sciences 84*, 141-145.

[47] Büll, C., Stoel, M. A., den Brok, M. H., and Adema, G. J. (2014) Sialic Acids Sweeten a Tumor's Life, *Cancer Research 74*, 3199-3204.

[48] Yu, S., Fan, J., Liu, L., Zhang, L., Wang, S., and Zhang, J. (2013) Caveolin-1 up-regulates integrin α2,6-sialylation to promote integrin α5β1-dependent hepatocarcinoma cell adhesion, *FEBS Letters 587*, 782-787.

[49] Seales, E. C., Jurado, G. A., Singhal, A., and Bellis, S. L. (2003) Ras oncogene directs expression of a differentially sialylated, functionally altered β1 integrin, *Oncogene 22*, 7137.

[50] Sakuma, K., Aoki, M., and Kannagi, R. (2012) Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial–mesenchymal transition, *Proceedings of the National Academy of Sciences 109*, 7776-7781.

[51] Zhuo, Y., Chammas, R., and Bellis, S. L. (2008) Sialylation of beta1 integrins blocks cell adhesion to galectin-3 and protects cells against galectin-3-induced apoptosis, *The Journal of biological chemistry 283*, 22177-22185.

[52] Swindall, A. F., and Bellis, S. L. (2011) Sialylation of the Fas death receptor by ST6Gal-I provides protection against Fas-mediated apoptosis in colon carcinoma cells, *The Journal of biological chemistry 286*, 22982-22990.

[53] Yamamoto, H., Oviedo, A., Sweeley, C., Saito, T., and Moskal, J. R. (2001) α2,6-Sialylation of Cell-Surface N-Glycans Inhibits Glioma Formation in Vivo, *Cancer Research 61*, 6822-6829.

[54] Sarrazin, S., Lamanna, W. C., and Esko, J. D. (2011) Heparan sulfate proteoglycans, In *Cold Spring Harb Perspect Biol*.

[55] Coppi, A., Tewari, R., Bishop, J. R., Bennett, B. L., Lawrence, R., Esko, J. D., Billker, O., and Sinnis, P. (2007) Heparan Sulfate Proteoglycans Provide a Signal to Plasmodium Sporozoites to Stop Migrating and Productively Invade Host Cells, *Cell Host & Microbe 2*, 316-327.

[56] Sanderson, R. D. (2001) Heparan sulfate proteoglycans in invasion and metastasis, *Seminars in Cell & Developmental Biology 12*, 89-98.

[57] Iozzo, R. V., and San Antonio, J. D. (2001) Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena, *The Journal of Clinical Investigation 108*, 349-355.

[58] Dreyfuss, J. L., Regatieri, C. V., Jarrouge, T. R., Cavalheiro, R. P., Sampaio, L. O., and Nader, H. B. (2009) Heparan sulfate proteoglycans: structure, protein interactions and cell signaling, *Anais da Academia Brasileira de Ciências 81*, 409-429.

[59] Nurcombe, V., Ford, M., Wildschut, J., and Bartlett, P. (1993) Developmental regulation of neural response to FGF-1 and FGF-2 by heparan sulfate proteoglycan, *Science 260*, 103-106.

[60] Cohen, T., Gitay-Goren, H., Sharon, R., Shibuya, M., Halaban, R., Levi, B. Z., and Neufeld, G. (1995) VEGF121, a vascular endothelial growth factor (VEGF) isoform lacking heparin binding ability, requires cell-surface heparan sulfates for efficient binding to the VEGF receptors of human melanoma cells, *J Biol Chem 270*, 11322-11326.

[61] Lortat-Jacob, H. (2009) The molecular basis and functional implications of chemokine interactions with heparan sulphate, *Current Opinion in Structural Biology 19*, 543-548.

[62] Tanaka, Y., Kimata, K., Adams, D. H., and Eto, S. (1998) Modulation of cytokine function by heparan sulfate proteoglycans: sophisticated models for the regulation of

cellular responses to cytokines, *Proc Assoc Am Physicians 110*, 118-125.

[63] Siegel, R. L., Miller, K. D., and Jemal, A. (2019) Cancer statistics, 2019, *CA: A Cancer Journal for Clinicians 69*, 7-34.

[64] Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation — potential for therapeutics and diagnostics, *Nature Reviews Drug Discovery 4*, 477.

[65] Zhang, S., Cordon-Cardo, C., Zhang, H. S., Reuter, V. E., Adluri, S., Hamilton, W. B., Lloyd, K. O., and Livingston, P. O. (1997) Selection of tumor antigens as targets for immune attack using immunohistochemistry: I. Focus on gangliosides, *International Journal of Cancer 73*, 42-49.

[66] Zhang, S., Zhang, H. S., Cordon-Cardo, C., Reuter, V. E., Singhal, A. K., Lloyd, K. O., and Livingston, P. O. (1997) Selection of tumor antigens as targets for immune attack using immunohistochemistry: II. Blood group-related antigens, *International Journal of Cancer 73*, 50-56.

[67] Liu, Y.-C., Yen, H.-Y., Chen, C.-Y., Chen, C.-H., Cheng, P.-F., Juan, Y.-H., Chen, C.-H., Khoo, K.-H., Yu, C.-J., Yang, P.-C., Hsu, T.-L., and Wong, C.-H. (2011) Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells, *Proceedings of the National Academy of Sciences 108*, 11332-11337.

[68] Hsu, T.-L., Hanson, S. R., Kishikawa, K., Wang, S.-K., Sawa, M., and Wong, C.-H. (2007) Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells, *Proceedings of the National Academy of Sciences 104*, 2614-2619.

[69] Vasseur, J. A., Goetz, J. A., Alley Jr, W. R., and Novotny, M. V. (2012) Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins, *Glycobiology 22*, 1684-1708.

[70] Arnold, J. N., Saldova, R., Galligan, M. C., Murphy, T. B., Mimura-Kimura, Y., Telford, J. E., Godwin, A. K., and Rudd, P. M. (2011) Novel glycan biomarkers for the detection of lung cancer, *Journal of proteome research 10*, 1755-1764.

[71] Ferdosi, S., Rehder, D. S., Maranian, P., Castle, E. P., Ho, T. H., Pass, H. I., Cramer, D. W., Anderson, K. S., Fu, L., Cole, D. E. C., Le, T., Wu, X., and Borges, C. R. (2018) Stage Dependence, Cell-Origin Independence, and Prognostic Capacity of Serum Glycan Fucosylation, β1–4 Branching, β1–6 Branching, and α2–6 Sialylation in Cancer, *Journal of Proteome Research 17*, 543-558.

[72] Ferdosi, S., Ho, T. H., Castle, E. P., Stanton, M. L., and Borges, C. R. (2018) Behavior of blood plasma glycan features in bladder cancer, *PloS one 13*, e0201208.

[73] Kang, P., Mechref, Y., and Novotny, M. V. (2008) High-throughput solid-phase permethylation of glycans prior to mass spectrometry, *Rapid Communications in Mass Spectrometry 22*, 721-734.

[74] Borges, C. R., Rehder, D. S., and Boffetta, P. (2013) Multiplexed surrogate

analysis of glycotransferase activity in whole biospecimens, *Analytical chemistry 85*, 2927-2936.

[75] Zaare, S., Aguilar, J. S., Hu, Y. M., Ferdosi, S., and Borges, C. R. (2016) Glycan Node Analysis: A Bottom-up Approach to Glycomics, *Jove-J Vis Exp*.

[76] Banazadeh, A., Veillon, L., Wooding, K. M., Zabet-Moghaddam, M., and Mechref, Y. (2017) Recent advances in mass spectrometric analysis of glycoproteins, *Electrophoresis 38*, 162-189.

[77] Ciucanu, I., and Costello, C. E. (2003) Elimination of oxidative degradation during the per-O-methylation of carbohydrates, *Journal of the American Chemical Society 125*, 16213-16219.

[78] Purdie, T., and Irvine, J. C. (1903) The alkylation of sugars, *Journal of the Chemical Society, Transactions 83*, 1021-1037.

[79] Haworth, W. N. (1915) A new method of preparing alkylated sugars, *Journal of the Chemical Society, Transactions 107*, 8-16.

[80] Hakomori, S.-I. (1964) A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide, *The Journal of Biochemistry 55*, 205-208.

[81] Sandford, P. A., and Conrad, H. (1966) The Structure of the Aerobacter aerogenes A3 (S1) Polysaccharide. I. A Reexamination Using Improved Procedures for Methylation Analysis*, *Biochemistry 5*, 1508-1517.

[82] Phillips, L. R., and Fraser, B. A. (1981) Methylation of carbohydrates with dimsyl potassium in dimethyl sulfoxide, *Carbohydrate Research 90*, 149-152.

[83] Blakeney, A. B., and Stone, B. A. (1985) Methylation of carbohydrates with lithium methylsulphinyl carbanion, *Carbohydrate Research 140*, 319-324.

[84] Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates, *Carbohydrate research 131*, 209-217.

[85] Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. (2005) Solid‐phase permethylation of glycans for mass spectrometric analysis, *Rapid communications in mass spectrometry 19*, 3421-3428.

[86] Ciucanu, I. (2006) Per-O-methylation reaction for structural analysis of carbohydrates by mass spectrometry, *Analytica chimica acta 576*, 147-155.

[87] Jay, A. (1996) The methylation reaction in carbohydrate analysis, *Journal of carbohydrate chemistry 15*, 897-923.

[88] Lippi, G., Guidi, G. C., Mattiuzzi, C., and Plebani, M. (2006) Preanalytical variability: the dark side of the moon in laboratory testing, *Clinical Chemistry and Laboratory Medicine (CCLM) 44*, 358-365.

[89] Stankovic, A. K. (2004) The laboratory is a key partner in assuring patient safety, *Clinics in laboratory medicine 24*, 1023-1035.

[90] (2005) Human plasma for fractionation, monograph 0853. Ph. Eur. Suppl 5.3 Strasbourg, France: Council of Europe.

[91] Bravo, M., Grancha, S., and Jorquera, J. (2006) Effect of temperature on plasma freezing under industrial conditions, *Pharmeuropa scientific notes 2006*, 31-35.

[92] Carrick, D. M., Mette, E., Hoyle, B., Rogers, S. D., Gillanders, E. M., Schully, S. D., and Mechanic, L. E. (2014) The use of biospecimens in population-based research: a review of the National Cancer Institute's Division of Cancer Control and Population Sciences grant portfolio, *Biopreservation and biobanking 12*, 240-245.

[93] Evans, M. J., Livesey, J. H., Ellis, M. J., and Yandle, T. G. (2001) Effect of anticoagulants and storage temperatures on stability of plasma and serum hormones, *Clinical biochemistry 34*, 107-112.

[94] Reisch, N., Reincke, M., and Bidlingmaier, M. (2007) Preanalytical stability of adrenocorticotropic hormone depends on time to centrifugation rather than temperature, *Clinical chemistry 53*, 358-359.

[95] Tanner, M., Kent, N., Smith, B., Fletcher, S., and Lewer, M. (2008) Stability of common biochemical analytes in serum gel tubes subjected to various storage temperatures and times pre-centrifugation, *Annals of clinical biochemistry 45*, 375-379.

[96] Ellervik, C., and Vaught, J. (2015) Preanalytical variables affecting the integrity of human biospecimens in biobanking, *Clinical chemistry 61*, 914-934.

[97] Holland, N. T., Pfleger, L., Berger, E., Ho, A., and Bastaki, M. (2005) Molecular epidemiology biomarkers—sample collection and processing considerations, *Toxicology and applied pharmacology 206*, 261-268.

[98] Kisand, K., Kerna, I., Kumm, J., Jonsson, H., and Tamm, A. (2011) Impact of cryopreservation on serum concentration of matrix metalloproteinases (MMP)-7, TIMP-1, vascular growth factors (VEGF) and VEGF-R2 in Biobank samples, *Clinical chemistry and laboratory medicine 49*, 229-235.

[99] Woodhams, B., Girardot, O., Blanco, M.-J., Colesse, G., and Gourmelin, Y. (2001) Stability of coagulation proteins in frozen plasma, *Blood coagulation & fibrinolysis 12*, 229-236.

[100] Rouy, D., Ernens, I., Jeanty, C., and Wagner, D. R. (2005) Plasma storage at− 80 C does not protect matrix metalloproteinase-9 from degradation, *Analytical biochemistry 338*, 294-298.

[101] Schiele, F., Vincent-Viry, M., Herbeth, B., Visvikis, A., and Siest, G. (2000) Effect of short-and long-term storage on human serum and recombinant apolipoprotein E concentration, *Clinical chemistry and laboratory medicine 38*, 525-528.

[102] Jeffs, J. W., Thibert, S. M., Ferdosi, S., Pham, L., Wilson, Z. T., Breburda, C.,

and Borges, C. R. (2019) Delta - S - Cys - Albumin: A Lab Test that Quantifies Cumulative Exposure of Archived Human Blood Plasma/Serum Samples to Thawed Conditions, *Molecular & Cellular Proteomics*, Submitted.

[103] Peters, T. (1995) The Albumin Molecule: Its Structure and Chemical Properties, In *All About Albumin* (Peters, T., Ed.), pp 9-75, Academic Press, San Diego.

[104] Naldi, M., Baldassarre, M., Domenicali, M., Bartolini, M., and Caraceni, P. (2017) Structural and functional integrity of human serum albumin: Analytical approaches and clinical relevance in patients with liver cirrhosis, *Journal of pharmaceutical and biomedical analysis 144*, 138-153.

[105] Arroyo, V., García-Martinez, R., and Salvatella, X. (2014) Human serum albumin, systemic inflammation, and cirrhosis, *Journal of Hepatology 61*, 396-407.

[106] Fernández, J., Monteagudo, J., Bargallo, X., Jiménez, W., Bosch, J., Arroyo, V., and Navasa, M. (2005) A randomized unblinded pilot study comparing albumin versus hydroxyethyl starch in spontaneous bacterial peritonitis, *Hepatology 42*, 627-634.

[107] Bertucci, C., and Domenici, E. (2002) Reversible and covalent binding of drugs to human serum albumin: methodological approaches and physiological relevance, *Current medicinal chemistry 9*, 1463-1481.

[108] Carballal, S., Radi, R., Kirk, M. C., Barnes, S., Freeman, B. A., and Alvarez, B. (2003) Sulfenic acid formation in human serum albumin by hydrogen peroxide and peroxynitrite, *Biochemistry 42*, 9906-9914.

[109] Nagumo, K., Tanaka, M., Chuang, V. T. G., Setoyama, H., Watanabe, H., Yamada, N., Kubota, K., Tanaka, M., Matsushita, K., and Yoshida, A. (2014) Cys34-cysteinylated human serum albumin is a sensitive plasma marker in oxidative stress-related chronic diseases, *PLoS One 9*, e85216.

[110] Roche, M., Rondeau, P., Singh, N. R., Tarnus, E., and Bourdon, E. (2008) The antioxidant properties of serum albumin, *FEBS letters 582*, 1783-1787.

[111] Taverna, M., Marie, A.-L., Mira, J.-P., and Guidet, B. (2013) Specific antioxidant properties of human serum albumin, *Annals of intensive care 3*, 4.

[112] Levine, R. L., Berlett, B. S., Moskovitz, J., Mosoni, L., and Stadtman, E. R. (1999) Methionine residues may protect proteins from critical oxidative damage, *Mechanisms of ageing and development 107*, 323-332.

[113] Lipinski, B. (2011) Hydroxyl radical and its scavengers in health and disease, *Oxidative medicine and cellular longevity 2011*.

[114] David, S., Balaram, P., and Mathan, V. (1995) Characterization of the interaction of lipid A and lipopolysaccharide with human serum albumin: implications for an endotoxin carrier function for albumin, *Journal of Endotoxin Research 2*, 99-106.

[115] Gioannini, T. L., Zhang, D., Teghanemt, A., and Weiss, J. P. (2002) An essential role for albumin in the interaction of endotoxin with lipopolysaccharide-binding protein

and sCD14 and resultant cell activation, *Journal of Biological Chemistry 277*, 47818-47825.

[116] Blomkalns, A. L., Stoll, L. L., Shaheen, W., Romig-Martin, S. A., Dickson, E. W., Weintraub, N. L., and Denning, G. M. (2011) Low level bacterial endotoxin activates two distinct signaling pathways in human peripheral blood mononuclear cells, *Journal of Inflammation 8*, 4.

[117] Pistolozzi, M., and Bertucci, C. (2008) Species‐dependent stereoselective drug binding to albumin: a circular dichroism study, *Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry 20*, 552-558.

[118] Matsuda, R., Anguizola, J., Joseph, K., and Hage, D. S. (2011) High-performance affinity chromatography and the analysis of drug interactions with modified proteins: binding of gliclazide with glycated human serum albumin, *Analytical and bioanalytical chemistry 401*, 2811.

[119] D Zhivkova, Z. (2015) Studies on drug–human serum albumin binding: the current state of the matter, *Current pharmaceutical design 21*, 1817-1830.

[120] Rael, L. T., Bar-Or, R., Ambruso, D. R., Mains, C. W., Slone, D. S., Craun, M. L., and Bar-Or, D. (2009) The effect of storage on the accumulation of oxidative biomarkers in donated packed red blood cells, *Journal of Trauma and Acute Care Surgery 66*, 76-81.

[121] Rehder, D. S., and Borges, C. R. (2010) Cysteine sulfenic acid as an intermediate in disulfide bond formation and nonenzymatic protein folding, *Biochemistry 49*, 7748-7755.

[122] Borges, C. R., Rehder, D. S., Jensen, S., Schaab, M. R., Sherma, N. D., Yassine, H., Nikolova, B., and Breburda, C. (2014) Elevated plasma albumin and apolipoprotein AI oxidation under suboptimal specimen storage conditions, *Molecular & Cellular Proteomics 13*, 1890-1899.

[123] Pecci, L., Montefoschi, G., Musci, G., and Cavallini, D. (1997) Novel findings on the copper catalysed oxidation of cysteine, *Amino acids 13*, 355-367.

[124] Kachur, A. V., Koch, C. J., and Biaglow, J. E. (1999) Mechanism of copper-catalyzed autoxidation of cysteine, *Free radical research 31*, 23-34.

[125] Anfinsen, C. B., and Haber, E. (1961) Studies on the reduction and re-formation of protein disulfide bonds, *J Biol Chem 236*, 1361-1363.

[126] Prevention, U. S. C. f. D. C. a. (2015-2016) National Health and Nutrition Examination Survey (NHANES).

[127] Jones, D. P., Mody Jr, V. C., Carlson, J. L., Lynn, M. J., and Sternberg Jr, P. (2002) Redox analysis of human plasma allows separation of pro-oxidant events of aging from decline in antioxidant defenses, *Free Radical Biology and Medicine 33*, 1290-1300.

[128] Blanco, R. A., Ziegler, T. R., Carlson, B. A., Cheng, P.-Y., Park, Y., Cotsonis, G. A., Accardi, C. J., and Jones, D. P. (2007) Diurnal variation in glutathione and cysteine redox states in human plasma, *The American journal of clinical nutrition 86*, 1016-1023.

[129] Stanley, P., Taniguchi, N., and Aebi, M. (2015-2017) N-glycans, In *Essentials of Glycobiology [Internet]*. (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[130] Brockhausen, I., and Stanley, P. (2015-2017) O-GalNAc Glycans, In *Essentials of Glycobiology [Internet]*. (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[131] Provencio, M., Isla, D., Sánchez, A., and Cantos, B. (2011) Inoperable stage III non-small cell lung cancer: Current treatment and role of vinorelbine, *Journal of thoracic disease 3*, 197.

[132] Team, N. L. S. T. R. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening, *New England Journal of Medicine 365*, 395-409.

[133] Okamura, K., Takayama, K., Izumi, M., Harada, T., Furuyama, K., and Nakanishi, Y. (2013) Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer, *Lung cancer 80*, 45-49.

[134] Xu, Y., Xu, L., Qiu, M., Wang, J., Zhou, Q., Xu, L., Wang, J., and Yin, R. (2015) Prognostic value of serum cytokeratin 19 fragments (Cyfra 21-1) in patients with non-small cell lung cancer, *Scientific reports 5*, 9444.

[135] Arrieta, O., Villarreal-Garza, C., Martínez-Barrera, L., Morales, M., Dorantes-Gallareta, Y., Peña-Curiel, O., Contreras-Reyes, S., Macedo-Pérez, E. O., and Alatorre-Alexander, J. (2013) Usefulness of serum carcinoembryonic antigen (CEA) in evaluating response to chemotherapy in patients with advanced non small-cell lung cancer: a prospective cohort study, *BMC cancer 13*, 254.

[136] Wang, X.-B., Li, J., and Han, Y. (2014) Prognostic significance of preoperative serum carcinoembryonic antigen in non-small cell lung cancer: a meta-analysis, *Tumor Biology 35*, 10105-10110.

[137] Bianchi, F., Nicassio, F., Marzi, M., Belloni, E., Dall'Olio, V., Bernard, L., Pelosi, G., Maisonneuve, P., Veronesi, G., and Di Fiore, P. P. (2011) A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer, *EMBO molecular medicine 3*, 495-503.

[138] Zheng, D., Haddadin, S., Wang, Y., Gu, L.-Q., Perry, M. C., Freter, C. E., and Wang, M. X. (2011) Plasma microRNAs as novel biomarkers for early detection of lung cancer, *International journal of clinical and experimental pathology 4*, 575.

[139] Shen, J., Todd, N. W., Zhang, H., Yu, L., Lingxiao, X., Mei, Y., Guarnera, M., Liao, J., Chou, A., and Lu, C. L. (2011) Plasma microRNAs as potential biomarkers for non-small-cell lung cancer, *Laboratory investigation 91*, 579.

[140] Belinsky, S. A., Klinge, D. M., Dekker, J. D., Smith, M. W., Bocklage, T. J., Gilliland, F. D., Crowell, R. E., Karp, D. D., Stidley, C. A., and Picchi, M. A. (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk, *Clinical Cancer Research 11*, 6505-6511.

[141] Balgkouranidou, I., Chimonidou, M., Milaki, G., Tsarouxa, E., Kakolyris, S., Welch, D., Georgoulias, V., and Lianidou, E. (2014) Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer, *British journal of cancer 110*, 2054.

[142] Hou, J.-M., Krebs, M., Ward, T., Sloane, R., Priest, L., Hughes, A., Clack, G., Ranson, M., Blackhall, F., and Dive, C. (2011) Circulating tumor cells as a window on metastasis biology in lung cancer, *The American journal of pathology 178*, 989-996.

[143] Ruhaak, L. R., Miyamoto, S., and Lebrilla, C. B. (2013) Developments in the identification of glycan biomarkers for the detection of cancer, *Molecular & Cellular Proteomics 12*, 846-855.

[144] Hu, Y., and Borges, C. R. (2017) A spin column-free approach to sodium hydroxide-based glycan permethylation, *Analyst 142*, 2748-2759.

[145] Gasperino, J. (2011) Gender is a risk factor for lung cancer, *Medical hypotheses 76*, 328-331.

[146] Osann, K. E., Anton‐Culver, H., Kurosaki, T., and Taylor, T. (1993) Sex differences in lung‐cancer risk associated with cigarette smoking, *International journal of cancer 54*, 44-48.

[147] Risch, H. A., Howe, G. R., Jain, M., Burch, J. D., Holowaty, E. J., and Miller, A. B. (1993) Are female smokers at higher risk for lung cancer than male smokers? A case-control analysis by histologic type, *American journal of epidemiology 138*, 281-293.

[148] Pope, M., Ashley, M., and Ferrence, R. (1999) The carcinogenic and toxic effects of tobacco smoke: are women particularly susceptible?, *The journal of gender-specific medicine: JGSM: the official journal of the Partnership for Women's Health at Columbia 2*, 45-51.

[149] Wakelee, H. A., Chang, E. T., Gomez, S. L., Keegan, T. H., Feskanich, D., Clarke, C. A., Holmberg, L., Yong, L. C., Kolonel, L. N., and Gould, M. K. (2007) Lung cancer incidence in never-smokers, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology 25*, 472.

[150] Stücker, I., Martin, D., Neri, M., Laurent-Puig, P., Blons, H., Antoine, M., Guiochon-Mantel, A., Brailly-Tabard, S., Canonico, M., and Wislez, M. (2017) Women Epidemiology Lung Cancer (WELCA) study: reproductive, hormonal,

occupational risk factors and biobank, *BMC public health 17*, 324.

[151] Wang, B.-Y., Huang, J.-Y., Cheng, C.-Y., Lin, C.-H., Ko, J.-L., and Liaw, Y.-P. (2013) Lung cancer and prognosis in Taiwan: a population-based cancer registry, *Journal of Thoracic Oncology 8*, 1128-1135.

[152] Hashimoto, S., Asao, T., Takahashi, J., Yagihashi, Y., Nishimura, T., Saniabadi, A. R., Poland, D. C., van Dijk, W., Kuwano, H., and Kochibe, N. (2004) α1-Acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis, *Cancer: Interdisciplinary International Journal of the American Cancer Society 101*, 2825-2836.

[153] Arnold, J. N., Saldova, R., Hamid, U. M. A., and Rudd, P. M. (2008) Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation, *Proteomics 8*, 3284-3293.

[154] Mizuguchi, S., Inoue, K., Iwata, T., Nishida, T., Izumi, N., Tsukioka, T., Nishiyama, N., Uenishi, T., and Suehiro, S. (2006) High serum concentrations of Sialyl Lewisx predict multilevel N2 disease in non–small-cell lung cancer, *Annals of surgical oncology 13*, 1010-1018.

[155] Mizuguchi, S., Nishiyama, N., Iwata, T., Nishida, T., Izumi, N., Tsukioka, T., Inoue, K., Kameyama, M., and Suehiro, S. (2007) Clinical value of serum cytokeratin 19 fragment and sialyl-Lewis x in non-small cell lung cancer, *The Annals of thoracic surgery 83*, 216-221.

[156] Mizuguchi, S., Nishiyama, N., Iwata, T., Nishida, T., Izumi, N., Tsukioka, T., Inoue, K., Uenishi, T., Wakasa, K., and Suehiro, S. (2007) Serum Sialyl Lewisx and cytokeratin 19 fragment as predictive factors for recurrence in patients with stage I non-small cell lung cancer, *Lung cancer 58*, 369-375.

[157] Iwata, T., Nishiyama, N., Nagano, K., Izumi, N., Tsukioka, T., Chung, K., Hanada, S., Inoue, K., Kaji, M., and Suehiro, S. (2012) Preoperative serum value of sialyl Lewis X predicts pathological nodal extension and survival in patients with surgically treated small cell lung cancer, *Journal of surgical oncology 105*, 818-824.

[158] Locher, C., Debieuvre, D., Coëtmeur, D., Goupil, F., Molinier, O., Collon, T., Dayen, C., Le Treut, J., Asselain, B., and Martin, F. (2013) Major changes in lung cancer over the last ten years in France: the KBP-CPHG studies, *Lung Cancer 81*, 32-38.

[159] Debieuvre, D., Oster, J.-P., Riou, R., Berruchon, J., Levy, A., Mathieu, J.-P., Dumont, P., Leroy-Terquem, E., Tizon-Couetil, V., and Martin, F. (2016) The new face of non-small-cell lung cancer in men: Results of two French prospective epidemiological studies conducted 10 years apart, *Lung cancer 91*, 1-6.

[160] Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease, *Cell 126*, 855-867.

[161] Varki, A. (2017) Biological roles of glycans, *Glycobiology 27*, 3-49.

[162] Dewald, J. H., Colomb, F., Bobowski-Gerard, M., Groux-Degroote, S., and Delannoy, P. (2016) Role of Cytokine-Induced Glycosylation Changes in Regulating Cell Interactions and Cell Signaling in Inflammatory Diseases and Cancer, *Cells 5*.

[163] Adua, E., Russell, A., Roberts, P., Wang, Y., Song, M., and Wang, W. (2017) Innovation Analysis on Postgenomic Biomarkers: Glycomics for Chronic Diseases, *Omics : a journal of integrative biology 21*, 183-196.

[164] Wada, Y., Azadi, P., Costello, C. E., Dell, A., Dwek, R. A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N. G., and Kato, K. (2007) Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study, *Glycobiology 17*, 411-422.

[165] Ciucanu, I. (2006) Per-O-methylation reaction for structural analysis of carbohydrates by mass spectrometry, *Anal Chim Acta 576*, 147-155.

[166] Hu, Y., Shihab, T., Zhou, S., Wooding, K., and Mechref, Y. (2016) LC–MS/MS of permethylated N‐glycans derived from model and human blood serum glycoproteins, *Electrophoresis 37*, 1498-1505.

[167] Zhou, S., Dong, X., Veillon, L., Huang, Y., and Mechref, Y. (2017) LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization, *Analytical and bioanalytical chemistry 409*, 453-466.

[168] Zhou, S., Wooding, K. M., and Mechref, Y. (2017) Analysis of Permethylated Glycan by Liquid Chromatography (LC) and Mass Spectrometry (MS), *Methods Mol Biol 1503*, 83-96.

[169] Goetz, J. A., Novotny, M. V., and Mechref, Y. (2009) Enzymatic/chemical release of O-glycans allowing MS analysis at high sensitivity, *Analytical chemistry 81*, 9546-9552.

[170] Mechref, Y., Kang, P., and Novotny, M. V. (2009) Solid-phase permethylation for glycomic analysis, In *Glycomics: Methods and Protocols* (Nicolle H, P., and Niclas G, K., Eds.), pp 53-64, Humana Press, New York City, NY.

[171] Harvey, D. J. (2011) Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry, *Journal of Chromatography B 879*, 1196-1225.

[172] Leymarie, N., and Zaia, J. (2012) Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis, *Analytical Chemistry 84*, 3040-3048.

[173] Powers, T. W., Jones, E. E., Betesh, L. R., Romano, P. R., Gao, P., Copland, J. A., Mehta, A. S., and Drake, R. R. (2013) Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked glycan expression in tissues, *Analytical chemistry 85*, 9799-9806.

[174] Tsai, T.-H., Wang, M., Di Poto, C., Hu, Y., Zhou, S., Zhao, Y., Varghese, R. S., Luo, Y., Tadesse, M. G., and Ziada, D. H. (2014) LC–MS profiling of N-glycans

derived from human serum samples for biomarker discovery in hepatocellular carcinoma, *Journal of proteome research 13*, 4859-4868.

[175] Skoog DA, W. D., Holler FJ, Crouch SR. (2014) Statistical Data Treatment and Evaluation, In *Fundamentals of analytical chemistry* 9th ed., pp 123-152, Cengage - Brooks/Cole, Belmont, CA.

[176] An, H. J., Peavy, T. R., Hedrick, J. L., and Lebrilla, C. B. (2003) Determination of N-glycosylation sites and site heterogeneity in glycoproteins, *Analytical chemistry 75*, 5628-5637.

[177] Moore, H. M., Compton, C. C., Lim, M. D., Vaught, J., Christiansen, K. N., and Alper, J. (2009) 2009 Biospecimen research network symposium: advancing cancer research through biospecimen science, AACR.

[178] Betsou, F., Gunter, E., Clements, J., DeSouza, Y., Goddard, K. A., Guadagni, F., Yan, W., Skubitz, A., Somiari, S., and Yeadon, T. (2013) Identification of evidence-based biospecimen quality-control tools: a report of the International Society for Biological and Environmental Repositories (ISBER) Biospecimen Science Working Group, *The Journal of Molecular Diagnostics 15*, 3-16.

[179] Carraro, P., Zago, T., and Plebani, M. (2012) Exploring the initial steps of the testing process: frequency and nature of pre-preanalytic errors, *Clinical chemistry 58*, 638-642.

[180] Borges, C., Jeffs, J. W., and Kapuruge, E. P. (2015) Impact of artifactual Ex Vivo oxidation on biochemical research, In *Oxidative Stress: Diagnostics, Prevention, and Therapy Volume 2*, pp 375-413, American Chemical Society.

[181] Engel, K. B., Vaught, J., and Moore, H. M. (2014) National Cancer Institute Biospecimen Evidence-Based Practices: a novel approach to pre-analytical standardization, *Biopreservation and biobanking 12*, 148-150.

[182] De Jongh, R., Vranken, J., Vundelinckx, G., Bosmans, E., Maes, M., and Heylen, R. (1997) The effects of anticoagulation and processing on assays of IL-6, sIL-6R, sIL-2R and soluble transferrin receptor, *Cytokine 9*, 696-701.

[183] Heins, M., Heil, W., and Withold, W. (1995) Storage of serum or whole blood samples? Effects of time and temperature on 22 serum analytes, *Clinical Chemistry and Laboratory Medicine 33*, 231-238.

[184] Lengellé, J., Panopoulos, E., and Betsou, F. (2008) Soluble CD40 ligand as a biomarker for storage-related preanalytic variations of human serum, *Cytokine 44*, 275-282.

[185] Karlsen, A., Blomhoff, R., and Gundersen, T. (2007) Stability of whole blood and plasma ascorbic acid, *European journal of clinical nutrition 61*, 1233.

[186] Lee, W., Roberts, S. M., and Labbe, R. F. (1997) Ascorbic acid determination with an automated enzymatic procedure, *Clinical chemistry 43*, 154-157.

[187] Ocké, M. C., Schrijver, J., Obermann-De Boer, G. L., Bloemberg, B. P., Haenen, G. R., and Kromhout, D. (1995) Stability of blood (pro) vitamins during four years of storage at− 20 C: consequences for epidemiologic research, *Journal of clinical epidemiology 48*, 1077-1085.

[188] Chaigneau, C., Cabioch, T., Beaumont, K., and Betsou, F. (2007) Serum biobank certification and the establishment of quality controls for biological fluids: examples of serum biomarker stability after temperature variation, *Clinical Chemical Laboratory Medicine 45*, 1390-1395.

[189] Turell, L., Carballal, S., Botti, H., Radi, R., and Alvarez, B. (2009) Oxidation of the albumin thiol to sulfenic acid and its implications in the intravascular compartment, *Brazilian journal of medical and biological research 42*, 305-311.

[190] Oettl, K., and Marsche, G. (2010) Redox state of human serum albumin in terms of cysteine-34 in health and disease, In *Methods in enzymology*, pp 181-195, Elsevier.

[191] Farrugia, A., Hill, R., Douglas, S., Karabagias, K., and Kleinig, A. (1992) Factor VIII/von Willebrand factor levels in plasma frozen to-30 C in air or halogenated hydrocarbons, *Thrombosis research 68*, 97-102.

[192] Sibinga, C. S., Das, P., and Meryman, H. T. (1990) *Cryopreservation and low temperature biology in blood transfusion*, Vol. 24, Springer Science & Business Media.

[193] Greilberger, J., Koidl, C., Greilberger, M., Lamprecht, M., Schroecksnadel, K., Leblhuber, F., Fuchs, D., and Oettl, K. (2008) Malondialdehyde, carbonyl proteins and albumin-disulphide as useful oxidative markers in mild cognitive impairment and Alzheimer's disease, *Free radical research 42*, 633-638.

[194] Terawaki, H., Yoshimura, K., Hasegawa, T., Matsuyama, Y., Negawa, T., Yamada, K., Matsushima, M., Nakayama, M., Hosoya, T., and Era, S. (2004) Oxidative stress is enhanced in correlation with renal dysfunction: examination with the redox state of albumin, *Kidney international 66*, 1988-1993.

[195] Bar‐Or, D., Heyborne, K. D., Bar‐Or, R., Rael, L. T., Winkler, J. V., and Navot, D. (2005) Cysteinylation of maternal plasma albumin and its association with intrauterine growth restriction, *Prenatal diagnosis 25*, 245-249.

[196] Borges, C. R., Oran, P. E., Buddi, S., Jarvis, J. W., Schaab, M. R., Rehder, D. S., Rogers, S. P., Taylor, T., and Nelson, R. W. (2011) Building multidimensional biomarker views of type 2 diabetes on the basis of protein microheterogeneity, *Clinical chemistry 57*, 719-728.

[197] van Eijsden, M., van der Wal, M. F., Hornstra, G., and Bonsel, G. J. (2005) Can whole-blood samples be stored over 24 hours without compromising stability of C-reactive protein, retinol, ferritin, folic acid, and fatty acids in epidemiologic research?, *Clinical Chemistry 51*, 230-232.

[198] Sokoll, L. J., Bruzek, D. J., Dua, R., Dunn, W., Mohr, P., Wallerson, G., Eisenberger, M., Partin, A. W., and Chan, D. W. (2002) Short-term stability of the

molecular forms of prostate-specific antigen and effect on percent complexed prostate-specific antigen and percent free prostate-specific antigen, *Urology 60*, 24-30.

[199] Chan, K. A., Yeung, S.-W., Lui, W.-B., Rainer, T. H., and Lo, Y. D. (2005) Effects of preanalytical factors on the molecular size of cell-free DNA in blood, *Clinical chemistry 51*, 781-784.

[200] Oddoze, C., Lombard, E., and Portugal, H. (2012) Stability study of 81 analytes in human whole blood, in serum and in plasma, *Clin Biochem 45*, 464-469.

[201] Woodrum, D., French, C., and Shamel, L. B. (1996) Stability of free prostate-specific antigen in serum samples under a variety of sample collection and sample storage conditions, *Urology 48*, 33-39.

[202] Kubasik, N. P., Graham, M., and Sine, H. E. (1979) Storage and stability of folate and vitamin B-12 in plasma and blood samples, *Clinica Chimica Acta 95*, 147-149.

[203] Evans, K., Mitcheson, J., and Laker, M. F. (1995) Effect of storage at 4 degrees C and-20 degrees C on lipid, lipoprotein, and apolipoprotein concentrations, *Clinical chemistry 41*, 392-396.

[204] Umemura, U., Koike, K. A., Sato, S., Iso, H., Shimamoto, T., and Komachi, Y. (1991) The effect of storage on serum fatty acids, *Nippon Eiseigaku Zasshi (Japanese Journal of Hygiene) 46*, 976-983.

[205] Qvist, P., Munk, M., Hoyle, N., and Christiansen, C. (2004) Serum and plasma fragments of C-telopeptides of type I collagen (CTX) are stable during storage at low temperatures for 3 years, *Clinica chimica acta 350*, 167-173.

[206] Boomsma, F., Alberts, G., Van Eijk, L., and Schalekamp, M. (1993) Optimal collection and storage conditions for catecholamine measurements in human plasma and urine, *Clinical chemistry 39*, 2503-2508.

[207] Cao, Z., Kamlage, B., Wagner-Golbs, A., Maisha, M., Sun, J., Schnackenberg, L., Pence, L., Schmitt, T., Daniels, J., Rogstad, S., Beger, R., and Yu, L.-R. (2019) An Integrated Analysis of Metabolites, Peptides and Inflammation Biomarkers for Assessment of Pre-analytical Variability of Human Plasma, *Journal of Proteome Research*, Under Review.

[208] Ruangturakit, S., Rojanasuphot, S., Srijunggravanvong, A., and Duangchanda, S. (1994) Storage stability of dengue IgM and IgG antibodies in whole blood and serum dried on filter paper strips detected by ELISA, *Southeast Asian Journal of Tropical Medicine and Public Health 25*, 560-560.

[209] Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J., Yi, J., Schuchard, M. D., Mehigh, R. J., Cockrill, S. L., Scott, G. B., and Tammen, H. (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples, *Proteomics 5*, 3262-3277.

[210] Lawrence, J. M., Umekubo, M. A., Chiu, V., and Petitti, D. B. (2000) Split sample analysis of serum folate levels after 18 days in frozen storage, *Clinical laboratory 46*,

483-486.

APPENDIX A

COPYRIGHT PERMISSIONS FOR ADAPTIONS OF FIGURES

Yueming Hu <yuemingh@asu.edu>

Your permission requests

1 message

Brown, Carol <brown@cshl.edu> Fri, Apr 5, 2019 at 12:43 PM
To: "yuemingh@asu.edu" <yuemingh@asu.edu>

Permission is granted for the use of Figs 6.1, 9.1, 10.1 and 11.1 from *Essentials of Glycobiology, Third Edition* in your PhD thesis only. Please include complete reference and copyright to Cold Spring Harbor Laboratory Press.

Best wishes for success with your thesis,

Carol C. Brown

Permissions Coordinator

Cold Spring Harbor Laboratory Press

500 Sunnyside Blvd

Woodbury, New York 11797

516 422 4038 ph.

516 422 4095 fx.

brown@cshl.edu

Molecular & Cellular Proteomics Copyright Policy

Parties who are not authors on the article

For noncommercial use:

Parties other than the authors seeking to reuse MCP content for noncommercial purposes are welcome to copy, distribute, transmit and adapt the work at no cost and without permission as long as they attribute the work to the original source using the citation above. Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations.
- Appending a reprinted article to a Ph.D. dissertation.

More details are provided in:

https://www.mcponline.org/node/48229/#copyright