

Medical University of South Carolina

**MEDICA**

---

MUSC Theses and Dissertations

---

2015

## Method Development for the Identification of Alternatively Sialylated Glycoproteins in Non-Small Cell Lung Cancer

Stephen Matthew Roper  
*Medical University of South Carolina*

Follow this and additional works at: <https://medica-musc.researchcommons.org/theses>

---

### Recommended Citation

Roper, Stephen Matthew, "Method Development for the Identification of Alternatively Sialylated Glycoproteins in Non-Small Cell Lung Cancer" (2015). *MUSC Theses and Dissertations*. 498.  
<https://medica-musc.researchcommons.org/theses/498>

This Dissertation is brought to you for free and open access by MEDICA. It has been accepted for inclusion in MUSC Theses and Dissertations by an authorized administrator of MEDICA. For more information, please contact [medica@musc.edu](mailto:medica@musc.edu).

Method Development for the Identification of Alternatively Sialylated  
Glycoproteins in Non-Small Cell Lung Cancer

by

Stephen Matthew Roper

A dissertation submitted to the faculty of the Medical University of South Carolina  
in partial fulfillment of the requirement for the degree of Doctor of Philosophy in  
the College of Graduate Studies

Department of Cell and Molecular Pharmacology and Experimental Therapeutics

2015

Approved by:

Chairman, Advisory Committee

Richard Drake

Lauren Ball

Robert Gemmill

Jennifer Isaacs

Michael Janech



# Abstract

Non-small cell lung cancer represents 85% of all lung cancers with an average 5 year life expectancy of 15-20%. A wealth of data suggests that altered glycosylation contributes to the progression of these tumors and efforts to improve the specificity of biomarkers have logically shifted towards glycoproteomic investigations. For this purpose, recent innovations in experimental strategies and analytical techniques could be combined to provide a more detailed characterization of glycans than previously achievable. These new methods have not yet been assessed for feasibility in experimental procedures useful for discovery phase efforts. Therefore, we aimed to investigate the utility of novel glycoproteomic and glycomic approaches for defining alterations in glycosylation which may accompany disease progression. To this end, we implemented an azido-sugar metabolic labeling, alkyne-agarose bead enrichment, and liquid chromatography-tandem mass spectrometry for profiling glycoproteins in a cell model of lung cancer induced to express transforming growth factor beta ligand-1. This approach identified putative changes in the sialylation of glycoproteins related to metabolic, cell adhesion, glycan biosynthesis, and extracellular matrix-related proteins. The application of a secondary digest to glycopeptide-bound beads, using peptide-N-glycosidase-F, was useful for verifying N-glycan sites as well as exposing previously undetected glycoproteins. A stable isotope labeling of amino acids in cell culture approach was used to gauge if altered protein expression was contributing to differential capture, however this method provided limited information due to a low overlap of proteins identified from enriched vs. unenriched fractions. In another set of experiments, we applied a novel derivatization strategy combined with high resolution/high mass accuracy mass spectrometry for discerning glycan structures in human lung cancer proximal fluids. This procedure effectively defined the sialic acid anomeric configuration of several prevalent species

and identified preliminary trends in the expression of oligomannose and complex glycans in clinically-relevant materials. Finally, matrix-assisted laser desorption ionization imaging mass spectrometry was used to spatially resolve the distribution of N-glycans in lung tissues matched to the proximal fluids. Histological assessment of these tissues facilitated cross-reference of acquired glycan species to regions of interest and provided a direct means for assessing how these trends correlated in the proximal fluids.

## TABLE OF CONTENTS

ABSTRACT .....	iii
LIST OF TABLES .....	viii
LIST of FIGURES .....	x
ACKNOWLEDGEMENT .....	xii
CHAPTERS	
1. Review of Literature .....	1
Overview .....	2
1.1 Non-small cell lung cancer .....	3
1.1.1 Introduction .....	3
1.1.2 Etiology and Detection .....	3
1.1.3 Sub-types and Features .....	6
1.2 Inflammation and TGF $\beta$ 1 in NSCLC .....	8
1.2.1 Inflammation and Pathogenesis of NSCLC .....	8
1.2.2 Introduction to TGF $\beta$ 1 .....	9
1.2.3 Canonical TGF $\beta$ 1 Signaling .....	10
1.2.4 Non-Canonical TGF $\beta$ 1 Signaling .....	11
1.2.5 Paradoxical Effects of TGF $\beta$ 1 in Cancer .....	12
1.2.6 TGF $\beta$ 1 and Invasive Tumor Phenotypes .....	13
1.2.7 TGF $\beta$ 1 and Glycosylation.....	16
1.3 Glycosylation and Biomarker discovery.....	18
1.3.1 Introduction to Glycoproteins.....	18
1.3.2 Biosynthesis and Roles of Glycoproteins .....	20
1.3.3 Sialic Acid.....	27
1.3.4 Glycosylation Alterations in Cancer .....	29
1.3.5 Sialic Acid in Cancer .....	35
1.3.6 Introduction to Glycoprotein Cancer Biomarkers.....	39
1.3.7 Current Glycoprotein Cancer Biomarkers .....	40
1.4 Mass spectrometry and Bioinformatics for Glycoproteomics .....	46
1.4.1 Introduction to MS-based Glycoproteomics.....	46
1.4.2 Enrichment Strategies for Glycoproteomics.....	47
1.4.3 Glycan Metabolic Labeling.....	50
1.4.4 Strategies for MS-based Glycoproteomics .....	53
1.4.5 Bioinformatics for Glycoproteomics .....	57
2. Hypothesis.....	60

3. Glycoproteomic and Glycomic Approaches to Evaluate TGFβ1-induced Effects on Sialoglycoproteins in a Cell Culture Model of NSCLC .....	65
3.1 Introduction.....	66
3.2 Materials and Methods.....	68
3.2.1 Cell Culture.....	68
3.2.2 Azido-sugar metabolic labeling.....	68
3.2.3 Agarose-alkyne Bead Enrichment of Azido-Sugar bearing Glycoproteins.....	70
3.2.3.1 Sample Preparation.....	70
3.2.3.2 Enrichment.....	72
3.2.4 Azido-sugar Incorporation Western Blot.....	72
3.2.5 LC-MS/MS and Bioinformatics for Label-Free Quantification .....	74
3.2.6 Double Metabolic Labeling.....	76
3.2.6.1 Sample Processing for Double Metabolic Labeling .....	76
3.2.7 LC-MS/MS and Bioinformatics for Double Metabolic Labeling.....	79
3.2.8 PNGase-F Mediated Release of N-linked Glycopeptides.....	81
3.2.9 LC-MS/MS and Bioinformatics for PNGase-F Mediated Release of N-linked Glycopeptides.....	82
3.2.10 H358 Global N-linked Glycomics .....	82
3.2.10.1 Linkage-Specific Sialic Acid Esterification.....	82
3.2.10.2 Permethylation .....	83
3.2.11 MALDI-FT-ICR N-Glycan Analysis and Bioinformatics.....	84
3.3 Results.....	85
3.3.1 Initial Quantification of Glycoproteins in TGFβ1, Zeb, and EV-Induced H358 Cells.....	85
3.3.2 Double Metabolic Labeling.....	90
3.3.3 PNGase-F release of N-linked glycopeptides.....	94
3.3.4 H358 Global N-linked Glycomics .....	96
3.4 Discussion.....	101
4. Glycomic Assessment of Human NSCLC Biofluids and Tissue.....	114
4.1 Introduction.....	115
4.2 Materials and Methods.....	117
4.2.1 Samples Description .....	117
4.2.2 Bronchoalveolar Lavage Fluid Proteomics.....	117
4.2.2.1 Sample Preparation and In-solution Digest .....	117
4.2.2.2 LC-MS/MS and Bioinformatics.....	118
4.2.3 N-linked Glycomics of Human Bio-fluids.....	118
4.2.4 N-linked Glycan Imaging of Lung Tissue .....	119
4.2.4.1 Sample Preparation .....	119
4.2.4.2 N-glycan MALDI Imaging Mass Spectrometry .....	119
4.3 Results.....	121
4.3.1 Bronchoalveolar Lavage Fluid Proteomics.....	121

4.3.2 N-linked Glycomics of Human Biofluids .....	123
4.3.3 N-linked Glycan Imaging of Lung Tissue .....	134
4.4 Discussion .....	139
5 Conclusions, Limitations, and Future Studies .....	155
5.1 Glycoproteomic and Glycomic Approaches to Evaluate TGF $\beta$ 1-induced Effects on Sialoglycoproteins in a Cell Culture Model of NSCLC .....	156
5.1.1 Conclusions.....	156
5.1.2 Limitations of Azido-sugar Metabolic Labeling Studies.....	158
5.1.3 Future Directions .....	160
5.2 Glycomic Assessment of Human NSCLC Biofluids and Tissue.....	164
5.2.1 Conclusions.....	164
5.2.2 Limitations of Human NSCLC Biofluids and Tissue Studies .....	165
5.2.2 Future Directions .....	167
REFERENCES .....	172
BIOGRAPHY .....	195



# List of Tables

<u>Table 1.</u> H358 Sample Groups for Azido-Sugar Metabolic Labeling and Enrichment, LC- MS/MS, and Label-Free Quantification .....	71
<u>Table 2.</u> H358 Sample Groups for Double Metabolic Labeling and Protein Quantification .... .....	77
<u>Table 3.</u> Individual Replicates of all Label-Free Quantification Experiments .....	Supplemental Material
<u>Table 4.</u> Summary of H358 Cell Lysate Label-Free Quantification Results .....	89
<u>Table 5.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: TGFβ1 vs. EV (+Dox +ManNAcAz) .....	Supplemental Material
<u>Table 6.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: +Dox +ManNAcAz vs. No Dox +ManNAcAz (TGFβ1) .....	Supplemental Material
<u>Table 7.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: TGFβ1 vs. Zeb (+Dox +ManNAcAz).....	Supplemental Material
<u>Table 8.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: TGFβ1 vs. EV (+Dox +GalNAcAz).....	Supplemental Material
<u>Table 9.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: +Dox +GalNAcAz vs. No Dox +GalNAcAz (TGFβ1) .....	Supplemental Material
<u>Table 10.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: Zeb vs. EV (+Dox +ManNAcAz).....	Supplemental Material
<u>Table 11.</u> Captured Protein Quantification and KEGG pathway analysis of H358 cell lysate: +Dox +ManNAcAz vs. No Dox +ManNAcAz (Zeb).....	Supplemental Material
<u>Table 12.</u> SILAC Only All Protein Quantified: TGFβ1 vs. EV (+Dox +ManNAcAz)..... .....	Supplemental Material
<u>Table 13.</u> SILAC+Click All Proteins Quantified: TGFβ1 vs. EV (+Dox +ManNAcAz) .....	Supplemental Material
<u>Table 14.</u> All Shared Proteins in SILAC-Click and SILAC-Only Fractions .....	Supplemental Material

<u>Table 15.</u> Shared Proteins Significantly Different in SILAC+Click Fraction and Not Significantly Different in SILAC-Only Fraction at 3 Benjamini-Hockberg P-value Cutoffs .....	92
<u>Table 16.</u> All Peptides Identified with Deamidated Asn Following PNGase-F Digest off of Alkyne Beads.....	Supplemental Material
<u>Table 17.</u> Selected List of Peptides Identified Following PNGase-F digest off of Alkyne Beads .....	Supplemental Material
<u>Table 18.</u> Proteins Identified from Alkyne Agarose Bead PNGase-F and Trypsin Digests.....	Supplemental Material
<u>Table 19.</u> Lung Cancer Bio-specimen Resource Network Sample Types.....	126
<u>Table 20.</u> Label-free Quantification of Individual Bronchoalveolar Lavage Samples.....	Supplemental Material
<u>Table 21.</u> Proteins Differentially Identified in Bronchoalveolar Lavage Cancer and Benign Groups.....	Supplemental Material
<u>Table 22.</u> Gene Ontology Analysis of Proteins Unique to Cancer Bronchoalveolar Lavage Fluid Samples.....	124

# List of Figures

<u>Figure 1.</u> Tetraacetylated Azido-Sugar Incorporation Onto Nascent Glycoproteins.....	52
<u>Figure 2.</u> Azido-Sugar Metabolic Labeling Workflow.....	69
<u>Figure 3.</u> Workflow of Click-Chemistry Enrichment and On-Bead Trypsin Digest.....	73
<u>Figure 4.</u> Rationale for Double Metabolic Labeling Approach .....	78
<u>Figure 5.</u> Cumulative Sample Workflow for H358 Azido-Sugar Metabolic Labeling, Enrichment and LC-MS/MS.....	86
<u>Figure 6.</u> H358 Cells Incorporate ManNAcAz.....	87
<u>Figure 7.</u> Principle of PNGase-F Release of N-linked Glycopeptides from Alkyne- Agarose Beads .....	95
<u>Figure 8.</u> H358 Permethylation +/- PNGase-F .....	98
<u>Figure 9.</u> H358 Permethylation TGF $\beta$ 1 vs. EV (m/z = 1150-5000).....	99
<u>Figure 10.</u> H358 Linkage-Specific Sialic Acid Esterification (m/z = 1150-2600).....	100
<u>Figure 11.</u> Summary of Azido-sugar Metabolic Labeling Workflows.....	113
<u>Figure 12.</u> Bronchoalveolar Lavage Fluid Protein Content.....	122
<u>Figure 13.</u> Bronchoalveolar Lavage Fluid Pool Total FT-ICR Spectra (m/z = 1200-3250)..... .....	125
<u>Figure 14.</u> Bronchoalveolar Lavage Fluid Pool FT-ICR Spectra (m/z =1790-2050).....	126
<u>Figure 15.</u> Bronchoalveolar Lavage Fluid FT-ICR Spectra (m/z = 2200-2480) .....	127
<u>Figure 16.</u> Bronchoalveolar Lavage Fluid FT-ICR Spectra (m/z = 2940-3185) .....	128
<u>Figure 17.</u> Saliva Pool Total FT-ICR (m/z = 1160-3250) .....	130
<u>Figure 18.</u> Saliva Pool FT-ICR Spectra (m/z = 1660-1900).....	131
<u>Figure 19.</u> Saliva Pool FT-ICR Spectra (m/z = 1925-2025).....	132
<u>Figure 20.</u> Saliva Pool FT-ICR Spectra (m/z = 2050- 2375).....	133
<u>Figure 21.</u> Workflow for Lung Tissue Imaging of N-glycans.....	135
<u>Figure 22.</u> Lung Tissue H&E with Pathology Notes .....	136
<u>Figure 23a.</u> Distribution of N-glycans in Lung Tissue by MALDI Imaging MS (m/z = 1250-1750) .....	137
<u>Figure 23b.</u> Distribution of N-glycans in Lung Tissue by MALDI Imaging MS (m/z = 1950-2700) .....	138

<u>Figure 24a.</u> Distribution of $m/z = 1955.7$ in Lung Tissue and Identification in BAL Pools .....	142
<u>Figure 24b.</u> Distribution of $m/z = 1955.7$ in Lung Tissue and Identification in Saliva Pools...	143
<u>Figure 25a.</u> Distribution of $m/z = 1485.6$ in Lung Tissue and Identification in BAL Pools .....	144
<u>Figure 25b.</u> Distribution of $m/z = 1485.6$ in Lung Tissue and Identification in Saliva Pools...	145
<u>Figure 26a.</u> Distribution of $m/z = 1419.5$ in Lung Tissue and Identification in BAL Pools .....	146
<u>Figure 26b.</u> Distribution of $m/z = 1419.5$ in Lung Tissue and Identification in Saliva Pools...	147
<u>Figure 27.</u> Distribution of $m/z = 1688.7$ in Lung Tissue and Identification in BAL Pools .....	148
<u>Figure 28.</u> Distribution of $m/z = 1905.7$ in Lung Tissue and Identification in BAL Pools .....	149
<u>Figure 29.</u> Summary of Workflow for NSCLC Biofluids and Tissue Imaging Glycomics .....	154
<u>Figure 30.</u> Bridging Findings Between In vitro and Human Samples Methods .....	171

# Acknowledgment

The work presented here would not have been possible without the guidance, care, and constant encouragement of several people. My parents, siblings, and extended family have been an unending source of love and inspiration throughout this journey. I am eternally grateful to have you all in my life and appreciate all of the support and kindness you have given me. I would also like to extend a sincere thank you to my mentor Dr. Richard Drake. I've benefitted greatly from your knowledge and experience, and have always had a good time in lab thanks to your positive attitude and good sense of humor. I appreciate all of the advice and guidance you have provided me throughout my graduate studies. Finally, I would like to thank my committee members: Drs. Lauren Ball, Robert Gemmill, Jennifer Isaacs, and Mike Janech. I sincerely appreciate all of the time and effort you have invested in me, thank you for sharing your scientific expertise with me.

# Chapter 1

## Literature Review

## Overview

Lung cancer accounted for the most cancer-related deaths in 2014, with 85% of these being classified as NSCLC. The late detection and heterogeneous progression of this disease plays a significant role in the mortality rate, therefore efforts to intervene are focused on the identification of biomarkers for early diagnosis and prognostic stratification. Here we present a review of the etiology, characteristics, and detection of NSCLC as a basic primer of disease pathology and why there is a need for additional metrics to predict outcomes. Next, TGF $\beta$ 1 signaling is outlined to highlight a pathway hypothesized to promote the acquisition of invasive tumor phenotypes. This information establishes the relevance of our experimental model of NSCLC, H358 cells induced to express TGF $\beta$ 1, as a system for screening for changes in the glycosylation of cell adhesion, extracellular matrix, and cytoskeletal proteins. We next provide an overview of glycoprotein biosynthesis, functional roles, and observed changes in glycans which often accompany transformation and metastasis. In addition, some of the mechanisms by which neoplastic cells produce specific glycan structures and the clinical utility for their identification are exemplified. These sections provide a base set of knowledge for understanding why cancer-associated glycosylation changes may occur and which structures are often modified. Furthermore, this information demonstrates the value of glycoproteomic endeavors by providing examples of how previous research has translated to clinical use. Finally, we present the methods utilized for mass spectrometry-based glycoproteomic investigations. Strategies and experimental approaches for the selective study of these molecules are compared and common acquisition techniques are

reviewed. Recent innovations in glycoproteomic mass spectrometry and data analysis, and the benefits they provide relative to previous methods, are featured.

## 1.1 Non-Small Cell Lung Cancer

### 1.1.1 Introduction

Lung cancer, including both non-small cell and small cell carcinoma, is responsible for the most cancer-related deaths in the United States. For 2014, the American Cancer Society estimates approximately 224,000 new cases of lung cancer and a mortality rate of about 159,000 for both sexes(1,2). Efforts introduced to reduce the risk for lung cancer development, such as smoking cessation therapy and early screening, have resulted in a decreasing mortality rate over the last 30 years for males and the last 10 years for females. While these trends are encouraging, the need for further intervention is evidenced by the outstanding difference in mortality between lung cancer and other prominent neoplasms like breast, prostate, and colon. In addition, the 5-year survival rate of lung cancer is among the lowest (17% for all stages) and the probability of identifying a tumor while it is still localized is dismal (15%)(2,3).

### 1.1.2 Etiology and Detection

Cigarette smoking is by far the leading cause of lung cancer. Numerous epidemiological studies implicate cigarette smoking in the development of lung disease and current estimates suggest it is responsible for about 80% of respiratory malignancies(1,2,4). Smokers are estimated to have a 20x increased probability of acquiring lung cancer of in their lifetime relative to never-smokers. Former smokers decrease their risk by about half of current smokers and the probability of developing a



tumor continues to decrease over time, albeit never returning to non-smoker levels. Further evidence of a causal link between smoking and lung cancer development is provided by the changes in the prevalence of histological subtypes over time accompanying public attitudes toward smoking; that is, as more people have quit smoking, there has been a shift in the occurrence of specific subtypes of lung cancer most associated with smoking(4). While smoking is clearly the most relevant environmental exposure in the etiology of lung cancer, it alone does not account for all cases. Other carcinogenic agents include asbestos, radiation, arsenic, chromium, uranium, nickel, polycyclic aromatic hydrocarbons, diesel exhaust, and radon ("extrinsic components")(4). From a pathological perspective, extrinsic factors insight an initial injury within the lung. As a normal biological response, inflammatory cells are recruited to these sites where they produce cytokines/chemokines, growth factors, and pro-angiogenic molecules to protect and induce healing of the affected tissue. Typically these responses are self-limiting and wane following repair of the underlying insult. However in the context of repeated injury, such as in a smoker, inflammatory and wound-healing programs persist, inducing increased cell proliferation(5–7). This increased cellular replication, in combination with exposure to carcinogenic agents, reactive oxygen species from immune infiltrate, and an altered microenvironment signaling milieu, promotes the accumulation of genetic mutations and may induce the initial transformation. Intrinsic components, such as inherent genetic variation and epigenetic modifications, may also play a role in the transition to cancer. For example, alterations in genes that control the metabolic conversion of pro-carcinogens to carcinogens, chromosomal instability, and promoter hypermethylation at certain loci, are all thought to contribute to pathogenesis(1,4,8).

The late detection of lung cancer plays a significant role in the high mortality rate. Signs and symptoms, including cough, hoarseness, bloody sputum, and weight loss, are often absent until the tumor has grown beyond a curable time-point. This is partially due to the formidable volume of the respiratory tract, allowing tumors to grow for some time before affecting other physiological processes. In addition, smokers may overlook signs/symptoms; attributing them as a consequence of the cigarettes themselves and not the underlying pathology. Efforts to reduce the mortality rate have therefore been focused on identifying tumors at earlier stages, when surgical intervention/chemotherapy may be curable(3,4,9). Imaging-based methods remain the gold-standard for the discovery of lung tumors, with low-dose spiral computed tomography (LDCT) being of particular utility(4). LDCT allows physicians to survey respiratory tissues for suspicious lesions in a manner more sensitive, faster, and with reasonable increases in radiation exposure relative to standard chest x-ray(9,10). Studies evaluating the use of LDCT as an early screening method, such as the National Lung Screening Trial (NLST), demonstrated its ability to detect cancer sooner than other methods and thereby decrease the mortality rate in a specific patient population(11). These results were promising enough to prompt the American Cancer Society to publish screening guidelines based on its use; however the study also unveiled a high rate of false-positive diagnoses associated with LDCT assessment alone. Subsequent research approaches aimed at increasing the specificity of screening are now focused on identifying molecular/proteomic biomarkers of early tumors in serum, proximal fluids and tissue, which could be used in combination with imaging for early diagnosis(9,12–15).

The histological and molecular attributes of lung cancer play an important role in prognostic and therapeutic considerations. Classifying lung cancer generally begins by grouping tumors into either non-small lung cancer (NSCLC) or small cell lung cancer (SCLC) subsets based on cytological characteristics of the primary tumor(s). NSCLC represents about 85% of all lung tumors, SCLC around 10%, and other rare forms (carcinoid) less than 5%(1). Within NSCLC, histologic sub-classifications of adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC) are made based on examinations of biopsies or cytological specimens; with immunohistochemistry being the most reliable means to differentiate sub-types(3,4,12,16). These sub-classifications were originally used to stratify patients into prognostic and therapeutic groups; however advances in molecular biology now permit for the identification of specific genetic mutations driving individual tumors. Consequently, therapeutic strategies have evolved for personalized medicine based on this information and a restructuring of the lung cancer classification system has been implemented to incorporate radiological, histological, and molecular typing(4,16,17).

### 1.1.3 Sub-types and Features

Adenocarcinomas represent the most prevalent subtype of NSCLC, accounting for about 50% of all new diagnoses. This classification encompasses cancers arising in secretory epithelial cells associated with glandular structures; which tend to form peripherally and grow slowly(4). ADCs develop in smokers more commonly than non-smokers; however, within the non-smoking population ADC is the predominant form. Within this classification (but not limited to), point mutations/overexpression in the epidermal growth factor receptor (EGFR) and rearrangement of the EML4-ALK gene

serve as predictive biomarkers for therapeutic efficacy of specific tyrosine kinase inhibitors (TKI)(4,8,18). In addition, acquired mutation in EGFR following TKI therapy indicates resistance to these drugs. Alterations in the K-Ras gene occur at a high frequency in lung adenocarcinomas and result in the constitutive activation of signaling pathways for proliferation and survival programs. Targeting this mechanism is an area of intensive research for clinical management purposes; unfortunately with minimal gains as of yet(8). Several other genetic aberrations have been documented, mediating effects in epigenetics, transcription, proteolysis, tumor suppressor and cell cycle control proteins(19). An important observation regarding these genetic anomalies is that they tend to occur in a mutually exclusive manner and often drive ADC development in non- or light smokers (4,8,18). Continued research efforts to uncover the molecular pathology of ADC will likely provide novel therapeutic targets as well as diagnostic/prognostic biomarkers for stratification of patients presenting with this subtype.

Squamous cell carcinomas frequently originate in the central portions of the respiratory tract; particularly in epithelial cells lining the bronchi. SCC tumorigenesis is highly associated with smoking and its incidence (40% of all NSCLC) has declined following public disclosure of the dangers of cigarettes(1). The pathology of SCC has been well-defined and follows a predictable progression in tissue dysplasia and genetic modification. Molecular changes accompanying this subtype include loss of heterozygosity in CDKNA, TP53, and RB1, as well as overexpression of genes SOX2 and TP63(19,20). Documentation of SCC driver mutations lags behind ADC, particularly because of sub-type prominence issues and tissue availability. More recently, SCC-associated alterations with potential therapeutic/prognostic value have been

identified in FGFR1, PI3KCA, and DDR2 genes. Conflicting information exists regarding the status of mutations associated with adenocarcinoma occurring in SCC; specifically EGFR, EML4-ALK, and K-Ras mutations. Some reports suggest that these particular events may be associated with SCC initiation /progression, while other research asserts that these findings were the result of mixed histologies in study specimens (i.e. adenosquamous carcinoma)(19,20). In fact, further definition of the molecular diversity in SCCs is currently an area of great interest. Importantly, this work will likely contribute to targeted therapies and prognostic indicators similar to the significance of findings in ADCs.

Large cell carcinoma is the least prevalent sub-type, representing about 10% of NSCLC cases. LCC usually begins in the periphery of the lung, grows rapidly and may secrete hormones (1,4). LCC is difficult to diagnose using traditional methodologies (i.e. immunohistochemistry) because of the inability to distinguish it from non-differentiated forms of ADC and SCC; therefore the LCC diagnosis commonly used when there is no clear ADC or SCC differentiation(17).

## 1.2 Inflammation and TGF $\beta$ 1 in NSCLC

### 1.2.1 Inflammation and the Pathogenesis of NSCLC

Persistent inflammation and chronic tissue remodeling, from environmental factors such as smoking, as well as contributions from intrinsic components, characterize the initiation of tumorigenesis in NSCLC. Inflammatory pathways are activated in response to the initial insult, leading to recruitment of immune cells and restructuring of the tissue by various cell populations. While this response is typically of a self-limiting

nature, chronic activation of these pathways not only propagates tumor formation, it is also hypothesized to advance existing cancer cells towards the acquisition of invasive phenotypes. Numerous cytokines, growth factors, and inflammatory mediators regulate these processes under tight temporal and spatial control. However, the dysregulated growth and signaling networks of tumor cells, in the presence of the inflammatory tumor microenvironment, may take advantage of these signals to propagate survival and metastasis. Amongst the altered signaling milieu present in inflammatory microenvironment, the pleiotropic cytokine transforming growth factor beta ligand 1 (TGF $\beta$ 1) is of particular interest for the promotion of invasive phenotypes.

### 1.2.2 Introduction to TGF $\beta$ 1

TGF $\beta$ 1 is a multi-functional cytokine contributing to physiological processes including tissue development, regulation of immunological responses, epithelial/stromal cell proliferation and differentiation, induction and down-regulation of inflammation, and the initiation and resolution of wound-healing(21,22). Three family members of TGF $\beta$  ligands exist (i.e. TGF $\beta$ 1, TGF $\beta$ 2, TGF $\beta$ 3). TGF $\beta$ 1 is the major form produced relevant to immune functions/wound-healing regulation, however all 3 ligands have potential to mediate similar signal transduction pathways because they may share common downstream 2nd messengers (21,23). The initial synthesis of TGF $\beta$ 1, in stromal or immune cell populations, results in the production of a pre-pro-protein form of the molecule, which requires proteolytic cleavage for activation. The latent precursor molecule is often secreted into the extracellular matrix, where it remains dormant until acted upon by molecules such as integrin  $\alpha_v\beta_6$ , matrix metalloproteases (MMPs), thrombospondin-1, and various other enzymes/proteases as biological context

dictates(21,24). Liberation of active TGF $\beta$ 1 from its regulatory molecules, latent TGF binding protein (LTBP) and latency associated peptide (LAP), results in binding of dimeric-ligand complexes to type II TGF $\beta$  receptor (TGF $\beta$ RII). This binding induces conformational changes on the homodimeric TGF $\beta$ RII, moving it into proximity to TGF $\beta$  receptor type I (TGF $\beta$ RI or ALK5) homodimers. Ligand-bound TGF $\beta$ RII forms a heterotetrameric unit with TGF $\beta$ RI, allowing cross phosphorylation of its (TGF $\beta$ RI) cytoplasmic tail via TGF $\beta$ RII's Ser/Thr kinase domain(25). At this point, the activated ligand/receptor complex is primed for the activation of downstream effectors capable of inducing canonical and non-canonical effects of TGF $\beta$ 1. An important consideration to note is that different TGF $\beta$ RI isoform subtypes (ALK1-7) may be preferential activated depending on cell lineage, resulting in activation of diverging downstream signaling effectors(22,23,25). Although the predominant TGF $\beta$ RI isoform activated in immune, epithelial, and fibroblast populations by TGF $\beta$ 1 is ALK5, recent research suggests some outcomes may also be mediated through ALK1 in endothelial and epithelial cells(23,25,26).

### 1.2.3 Canonical TGF $\beta$ 1 Signaling

The canonical pathway initiated by ligand-bound TGF $\beta$ RII-ALK5 complexes begins with the phosphorylation of receptor-Smad proteins (R-Smads: Smad-2 and -3) by type I TGF $\beta$ RI, a Ser/Thr kinase. R-Smads are brought into proximity to activated receptor complexes by Smad-anchor for receptor activation (SARA), allowing each to be phosphorylated. Following this activation, Smad-2 and Smad-3 individually form heterodimers with the common Smad (Smad-4) and translocate to the nucleus where they function as transcription factors(22,25,27). Smad-binding elements (SBE) on pertinent

genes are targeted for modulation by a combination of R-Smad/Smad-4 and various co-activators or co-repressor molecules. Importantly, these co-regulatory molecules, such as AP-1, Ets, basic helix loop helix, Forkhead and Homeobox-containing transcription factor families, confer tight binding and specificity to genes regulated by TGF $\beta$ 1(22,28). R-Smads also recruit histone acetyltransferases (CBP/p300) and de-acetylases (Sno/Ski) in their complex orchestration of transcription control. Depending on the biological context/cell type, different co-regulatory molecules may preferentially interact with the R-Smads to influence gene expression. Some of transcription targets of TGF $\beta$ 1 relevant to inflammation/wound-healing include genes affecting morphology ( $\alpha$ -actinin, tropomyosin), extracellular matrix-modifying enzymes (collagens, MMPs, plasminogen activator inhibitor), and cell adhesion properties (cadherins, integrins, fibronectin)(29,30). In addition, canonical TGF $\beta$ 1 signaling results in the transcription of Smad-7 (inhibitory-Smad). Through this mechanism TGF $\beta$ 1 is able to regulate its own signaling via negative feedback; by direct competition of Smad-7 with Smad-2/3 for binding at TGF $\beta$ RI as well as the recruitment of an ubiquitin ligase for degradation of ALK5.

#### 1.2.4 Non-Canonical TGF $\beta$ 1 Signaling

Several studies have also uncovered the ability of TGF $\beta$ 1-bound TGF $\beta$ RII/TGF $\beta$ RI to affect cell-signaling pathways that are Smad-independent or that interact with Smads outside of the mechanism described above ("non-canonical")(22,27,31). Models of TGF $\beta$ 1 signaling using mutated TGF $\beta$ RI or Smad-4 deficiency demonstrate activation of pathways such as p38 MAPK, JNK, and Erk MAPK following stimulation(31). Other effectors such as PI3K-Akt and NF- $\kappa$ B are reportedly



stimulated by kinase cascades following TGF $\beta$ 1 ligand-binding; however the distinct means by which this is achieved has yet to be elucidated. One hypothesis is that TGF $\beta$ -induced kinase (TAK1) and Ras propagate transduction through these pathways via direct TGF $\beta$ 1 activation(27). An important outcome of non-canonical activation of these pathways is convergence with Smad-dependent signaling. TGF $\beta$ 1 activation of MAPK signaling has been shown to regulate Smad-dependent transcription targets through effects on co-regulatory transcription factors(22). The potential results mediated by the interaction of these pathways are diverse and confer a further level of regulation for generating context-dependent outcomes in TGF $\beta$ 1 signaling. Other effectors in non-canonical TGF $\beta$ 1 signaling include the Rho-like GTPases (RhoA, Rac, Cdc42) and PP2A(27,31). These molecules impart effects ranging from the modification of cytoskeletal organization, to serving as intermediates for the regulation of other non-canonical pathway molecules, as well as affecting cell cycle progression. Detailed characterization of Smad-independent signaling, and its significance to physiological responses of TGF $\beta$ 1 is still underway.

### 1.2.5 Paradoxical Effects of TGF $\beta$ 1 in Cancer

TGF $\beta$ 1 is well-known to promote growth-inhibition on non-transformed epithelial cells via canonical signaling; through modification of specific transcriptional targets controlling cell cycle and apoptosis(28). In the context of carcinogenesis, these actions suggest that the molecule promotes the suppression of tumor growth. In support of this role, retrospective analyses have identified mutated forms of TGF $\beta$ 1 receptors/effectors present in tumors of various origins; inferring that altered signaling may contribute to initial tumor formation(23,32). These functions of TGF $\beta$ 1 became particularly perplexing

following the observation that many advanced tumors over-express the ligand, and, that loss of growth-inhibitory response increased metastases and resulted in poor prognoses(25,33–36). Subsequent investigations, aimed at clarifying these apparently paradoxical roles for TGF $\beta$  as a tumor suppressor and promoter, were carried out using transgenic murine models of cancer with various modified components of the TGF $\beta$ 1 pathway. The studies confirmed the dual functionality of the cytokine and provided insight into some of the mechanisms established tumors use to overcome growth-inhibitory effects of TGF $\beta$ 1 and utilize them for the propagation of survival/invasion.

### 1.2.6 TGF $\beta$ 1 and Invasive Tumor Phenotypes

TGF $\beta$ 1 is initially secreted into inflammatory tumor microenvironments by immune infiltrate and stromal cells recruited to protect/repair underlying tissue damage. However, over time, transformed epithelial cells acquire the ability to produce and secrete the cytokine (37). In addition, latent TGF $\beta$ 1 present in the extracellular matrix may be activated by the enzymes/proteases secreted by these cells during the process of tissue restructuring. The effects of active TGF $\beta$ 1 within this physiological context influence the genetic programs of tumor and stromal cells, as well as regulating host-tumor immune interactions and promoting angiogenesis. One such effect, hypothesized to play a central role in the acquisition of invasive phenotypes, involves modification of a set of genes controlling epithelial cell differentiation, morphology, and function. The phenotypic changes incurred as a result of targeting these genes are collectively referred to as the epithelial to mesenchymal transition (EMT). EMT is a normal biological occurrence during embryogenesis, tissue development, and wound-healing. It provides a reversible mechanism for differentiated epithelial cells to undergo phenotypic transition

into de-differentiated, migratory states (mesenchymal) for the colonization of tissues in processes like organ development or wound repair(38). This program is typically under tight regulatory control; requiring specific extracellular cues and the cooperation of several molecular effectors for coordination. However, in a state of altered signaling, such as chronic inflammation and/or carcinogenesis, the inappropriate activation of EMT may occur and propagate pathological progression(28). An example of this is evidenced in states of organ fibrosis. In such cases, habitual EMT, stemming from un-resolved inflammation, TGF $\beta$ 1 production, and other molecular contributors, results in increased epithelial cell de-differentiation to fibroblast precursors(39). These fibroblast-like cells produce collagen in an attempt restructure the supporting tissue. Without resolution of the EMT-promoting events, hardening and dysfunction of the affected organ may result(40).

More recently, EMT has been hypothesized to be a mechanism for initiating metastasis in epithelial based tumors. Murine models and In vitro experiments demonstrate the ability of tumorigenic epithelial cells to respond to EMT-inducing molecules and acquire phenotypic changes consistent with mesenchymal cells. These phenotypic changes include loss of epithelial cell markers E-cadherin, Zona Occludens-1, and MUC1. In addition, these cells gain markers such as  $\alpha$ -smooth muscle actin, fibroblast-specific protein, N-cadherin, and vimentin; all highly suggestive of a mesenchymal-like state(28,38). TGF $\beta$ 1, along with other growth factors from within the tumor microenvironment, has been shown to induce the genetic changes consistent with acquisition of these phenotypic alterations; for example, through the modification of genes such as Snail, Slug, Zeb-1/2, Goosecoid, and Twist. TGF $\beta$ 1 imparts its pro-EMT effect through both canonical and non-canonical pathways, with cooperative interaction

of certain pathway effectors resulting in a synergistic effect. Canonical signaling influences the expression of target genes mentioned above, through direct activation/repression of transcription factors or epigenetic modifications. Smad-independent signaling, such as through Ras activation, promotes EMT by modifying adherens junctions as well as working together with canonical signals for the induction of genes like Snail. In addition, non-canonical signaling contributes the cytoskeletal organization changes and the dissolution of tight junctions, through interactions with molecules like Erk, JNK/p38, Rho-like GTPases, and CDC42. Each pathway is also hypothesized to contribute to the autocrine production of TGF $\beta$ 1 by transformed cells, feeding back to strengthen these signaling outcomes(28,38,41). Furthermore, oncogenic changes within the cells may assist in TGF $\beta$ 1-induced EMT. For example, loss of the p53 tumor suppressor, in combination with TGF $\beta$ 1 signaling, promotes the de-regulation of key mediators of cell cycle progression(42). Outside of the direct tumor cell effects of TGF $\beta$ 1, the cytokine also primes the tumor microenvironment for invasion. Immune cells, such as neutrophils and monocytes, are recruited to the tumor by TGF $\beta$ 1 and other inflammatory mediators. Upon arrival, these cells secrete matrix-modifying enzymes, growth factors and cytokines as a normal physiological response(21,43). The dissolution of the extracellular matrix by matrix-metalloproteases, collagenases, and other proteases by these cells may help facilitate intravasation of tumor cells as an initial step in metastasis. Finally, TGF $\beta$ 1 skews the phenotype of immune cells away from anti-tumor responses, further supporting aggressive tumor growth(24).

### 1.2.7 TGFβ1 and Glycosylation

Several of the proteins targeted for modification by TGFβ1 in the acquisition of invasive tumor phenotypes share functional characteristics: being related to cell adhesion, cytoskeletal organization, and extracellular matrix (ECM) remodeling. The change in differentiation state of tumor cells stimulated with TGFβ1 requires a comprehensive re-organization of these types of molecules and their biological properties. Beyond effects on protein expression levels, other characteristics likely regulated by TGFβ1 influence protein-protein interaction, localization and stability. Based on this rationale, investigations are now exploring additional mechanisms by which TGFβ1 influences tumor cell phenotypes. For example, several studies are investigating alterations in protein glycosylation as a potential means to propagate TGFβ1-induced invasive characteristics(44–47). The principle behind this is seeded in the abundance of relevant biological functions that glycans carry out and the high frequency with which these PTM's are attached to proteins modified during invasion. Glycans influence cell-cell and cell matrix-adhesion, protein-trafficking, ligand-receptor interaction, and immune recognition(48). Furthermore, a wealth of literature demonstrates alterations in glycosylation profiles accompanying embryonic tissue development and cell differentiation, neoplastic transformation, and metastatic dissemination(48–53).

Examinations of microarray data from cell models of cancer stimulated with TGFβ1 have uncovered significant differences in the mRNA levels of potentially relevant enzymes responsible for glycan synthesis, extension, and degradation. For example, Maupin et al. detected alterations in the levels sialyl-transferases and polypeptide N-acetylgalactosaminyltransferases transcripts following TGFβ1 stimulation(45). This

group further demonstrated that various pancreatic cancer cell lines, designated as either epithelial or mesenchymal, present with vast differences in expression profiles of genes vital to branching/lengthening of existing glycans and O-type glycosylation. Another complementary analysis revealed contrasts in glycan-related gene expression between actively migrating and stationary cancer cells. Conclusions drawn from their efforts indicate that glycosylation-related genes, as a group, are significantly enriched for modification following TGF $\beta$ 1 stimulation. However, clear relationships regarding how TGF $\beta$ 1 influences these genes have yet to be elucidated. Zhang et al. provide evidence of a pSmad-2 dependent effect for the induction of the glycosyl transferase *GCNT2* and demonstrate the requirement of its enzymatic activity in the induction of certain migratory characteristics(46). Other suspected pathways may be influenced by TGF $\beta$ 1's ability to affect epigenetic modifications. The *MGAT3* gene codes for the glycosyl transferase GnT-III, which catalyzes the addition of bisecting N-acetylglucosamine (GlcNAc) to core N-glycan structures. The presence of these bisecting structures precludes further branching of glycan chains catalyzed by the glycosyl transferase (GnT-V). An important observation has been that increased levels of GnT-V, and decreased levels of GnT-III, are associated with increased metastases in some cancer models(54–56). Pinho et al. used this information to research mechanisms by which aggressive cancers may modify the expression of these enzymes(47). Their work established that promoter hypermethylation of *MGAT3* occurs following TGF $\beta$ 1 exposure and they conclude that the subsequent loss of bisecting GlcNAc on E-cadherin may allow for an increased turnover rate of the cell adhesion molecule. Beyond cell-specific effects, other studies have put forth hypotheses proposing that TGF $\beta$ 1 can modify the glycosylation

status of vital ECM components in contributing to invasion. Freire-de-Lima et al. revealed increased enzymatic activity of specific N-acetylgalactosaminyl transferases (GalNAc-T) following TGF $\beta$ 1 stimulation of prostate cancer cell lines(57). These GalNAc-Ts were shown to be responsible for the addition of a specific O-linked glycan on fibronectin, which defines a variant form of the protein associated with metastatic cancer (oncofetal fibronectin). The authors conclude their study by demonstrating that blocking oncofetal fibronectin expression, through upstream inhibition of the GalNAc-Ts, attenuates TGF $\beta$ 1-mediated acquisition of invasive phenotypes in these cells.

## 1.3 Glycosylation and Biomarker Discovery

### 1.3.1 Introduction to Glycoproteins

Glycosylation is the most common post translational modification, estimated to occur on around 50% of all mammalian proteins. These sugar molecules ("glycans") carry out various biological functions, which can be broadly defined into structural/modulatory roles and the regulation of cell-cell and cell-environment recognition events. Glycosylation may occur on proteins, lipids, DNA, and other organic molecules; with the collective group of substrates being referred to as glycoconjugates. For proteins in particular, the attachment of carbohydrate molecules influences conformation, localization, retention, and protein-protein interaction. At the cellular level, this translates into effects on processes such as cell-cell and cell-matrix adhesion, receptor-ligand binding, ECM structure and integrity, and cytoskeletal dynamics (58,59). Consequently, glycans are often found attached to secreted or cell-membrane associated proteins, although cytosolic and nuclear glycoproteins also exist. From an evolutionary

standpoint, glycosylation is thought to increase the diversity of functions that a limited set of molecules (proteins, lipids) may carry out for normal biological function and in responding to environmental stimuli. Supporting this idea is the array of mono- and polysaccharide molecules used to synthesize these structures. Carbohydrates may be attached in alternative orders, differing anomeric configurations, and in branched or linear configurations for a given glycosite on a single protein under different physiological conditions (48). In addition, glycans may be further modified by sulfation, methylation, or acetylation following attachment. The relevance of glycosylation in influencing biological processes is evident in the multitude of disorders identified from its dysregulation. For example, glycans play a role in pathogenic, congenital, and acquired disease states. The congenital disorders of glycosylation (CGD) are a collection of inherited maladies caused by ineffective glycan anabolism commonly presenting at early stages of life. In addition, several of the lysosomal storage syndromes, including Tay-Sachs disease, are related to the inability to properly degrade glycans(48,60,61). A common theme in these glycosylation disorders is that pathology is often evident at early stages of life, reflecting the nature of these molecules to carry out functions in developmental/differentiation cell programs(62). In line with this were observations that cancer cells present with an altered repertoire of glycans relative to non-transformed cells of the same origin. Studies using lectins (molecules which recognize glycan epitopes), glycan-specific antibodies, and protein separation techniques initially uncovered the changing "glycome" tumors often exhibit(63). Subsequent efforts confirmed that the repertoire was again changing as tumor cells undergo metastatic dissemination and



hypotheses began being formed regarding how tumor cells may utilize these changes to propagate the neoplastic phenotype.

### 1.3.2 Biosynthesis and Roles of Glycoproteins

The potential for great diversity in glycan structures attached to proteins in mammalian cells exist. At present, 10 different precursor sugar molecules are known, which are derived from metabolic and/or glycan salvage pathways(64). These monosaccharide molecules are "activated" in the cytoplasm by attachment to high-energy nucleotide-phosphates and then transported to the endoplasmic reticulum and Golgi apparatus. Within these compartments, the molecules now serve as donor-substrates for the enzymes that initiate and elongate these structures. The glycosyltransferases (GTs), which catalyze the formation of substrate-glycan bonds, typically act in a manner consistent with one enzyme being responsible for the attachment of one donor sugar to an acceptor substrate in a specified linkage conformation(65). However, within this large group of highly conserved enzymes (over 96 GT families exist), exceptions to the "one enzyme-one linkage" observation exist. For example, cases where more than one GT catalyzes the addition of the same donor-substrate and linkage conformation have been demonstrated (i.e. shared specificity). Other exceptions include single GTs being capable of adding a donor substrate in different linkages or being promiscuous in acceptor specificity(65). GTs may also compete amongst themselves for donor glycan substrates, and as well, cell/tissue specific expression of different GT families has been observed(64). In addition, the enzymes which hydrolyze glycan-substrate linkages, the glycosidases, function along the anabolic pathway creating intermediate glycan structures. These exceptions to the one enzyme-one linkage observation have confounded efforts to predict

glycan structures based on the presence/absence of specific GTs in different cell types. However, other useful information regarding these enzymes has been deduced with potential applications for uncovering specific glycan roles(66). For instance, the initiating-GTs tend to have higher levels of specificity for acceptor substrates (polypeptides, lipids, nucleic acids) relative to later-acting GTs which extend oligosaccharides (where acceptor substrates are previously attached glycans). Consensus sequence/motifs have been identified for some of the initiating enzymes, aiding in the calculation of where these structures may be attached. Furthermore, this information was useful for elucidating the molecular mechanisms by which carbohydrates are bound to other molecules. As an example, protein N-linked glycosylation only occurs in the amino acid consensus sequence Asn-X-Ser/Thr, where X is any amino acid other than Proline. The biological rationale for this consensus sequence is evidenced by amide bond formation between the initial GlcNAc molecule of the glycan and the nitrogen on the side chain of Asn. It should be noted for this particular example, however, that just because a consensus sequence is present does not necessarily mean glycosylation will be initiated; the same being true for glycan acceptor substrates of the extending GTs. Further complexity in carbohydrate structures are mediated through monosaccharide substrate availability, the potential for multiple anomeric configurations, and chemical modification of glycans(64,67). These examples provide only a partial insight into some of the mechanisms responsible for the exuberant amount of oligosaccharide diversity in mammalian cells. Importantly, cells utilize these assortments of glycan structures/attachments to influence the functionality of glycoconjugates and may alter the structures in response to different biological contexts(67).

Three major types of mammalian glycoproteins exist: N-linked, O-linked, and proteoglycans. As mentioned above, N-linked glycosylation is defined by bond formation between Asn residues in the consensus sequence Asn-X-Ser/Thr and a GlcNAc substrate. O-linked glycans are attached ambiguously at Ser/Thr, through a reaction with the side chain hydroxyl group of these amino acids. The third class of glycoproteins, proteoglycans, is composed of a protein backbone with one or more glycosaminoglycan (GAG) branches occurring along the length of the core. These linear, highly-sulfated molecules are abundant in the ECM where they influence biological processes through charge state effects from the GAG molecules they carry(68). N- and O- glycans share some characteristics in their biosynthesis and function. Each of these molecules exerts distinct influences on the underlying protein to which they are attached based on the unique repertoire of sugars they carry.

The synthesis of N-linked glycans begins in the ER and is completed in the Golgi, where structural complexity may be added by concerted efforts of glycosyltransferases and glycosidases. The glycoprotein products are typically destined for secretion or insertion into the cell membrane; a fate that is inherently tied to the localization of enzymes required for their production (66). The N-linked glycan pathway is unique in that these molecules are first constructed by the transfer of a "core" set of glycans from a dolichol-phosphate donor. This highly conserved, constitutive pathway is orchestrated co-translationally, as a growing polypeptide chain enters the ER. First, a 14-sugar core ( $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ ) is covalently linked to the substrate by the oligosaccharide complex (OST), yielding an immature N-glycan. Glycosidases present in the ER and Golgi cleave glucose and mannose residues from the core, typically resulting in N-glycan structures of

Man<sub>5</sub>GlcNAc<sub>2</sub>(69). These "fully trimmed" glycans may now be elongated with various monosaccharides through the action of medial and trans- Golgi-based GTs. However, not all N-linked glycans undergo full trimming, and as well, not all trimmed N-glycans will be elongated. The carbohydrate moieties present on N-linked glycoproteins which retain >5 core mannose sugars cannot be further processed by Golgi GTs. This type of N-glycan is referred to as "oligomannose" and variations in their presentation on final protein products range from fully trimmed (Man<sub>5</sub>GlcNAc<sub>2</sub>) to untrimmed (Man<sub>9</sub>GlcNAc<sub>2</sub>) and anywhere in-between(69). Another important biological consequence of the trimming reactions is the facilitation of protein quality control, through interactions with ER-based chaperones. Unfolded or mis-folded proteins may not be appropriately processed by the glycosidases responsible for removing core glucose molecules, resulting in retention and possible proteosomal degradation of the polypeptide(70). Those glycans which successfully undergo complete trimming have the additional potential to become either "complex" or "hybrid" type N-glycans(69). Complex N-glycans are generated from the Man<sub>5</sub>GlcNAc<sub>2</sub> core structures through reactions mediated by GlcNAc-transferase I & II as well as the glycosidase  $\alpha$ -mannosidase II. The action of these enzymes further reduces the mannose content ( to Man<sub>3</sub>GlcNAc<sub>2</sub>) and adds GlcNAc residues to both of the core "branches", providing an acceptor substrate for further GT action. In addition each branch may form a bond with up to two GlcNAc linkages, resulting in bi-, tri-, and tetra-antennary and bisecting complex glycans. Hybrid N-glycans, as the name implies, are a combination of oligomannose and complex structures. This type of N-glycan is produced when  $\alpha$ -mannosidase II cleaves only 1 mannose from the fully trimmed core, resulting in GlcNAc<sub>1</sub>Man<sub>4</sub>GlcNAc<sub>2</sub>. The terminal GlcNAc residue present on one of the branches

may be extended with other sugars (like complex) while the other branch is terminated with core-derived mannose (like oligomannose). As complex and hybrid N-glycans progress through the Golgi, various modifications may occur to the core and GlcNAc branches. For example, the addition of a fucose molecule to the GlcNAc bound to Asn ("core fucosylation") or glycosidic linkages between GlcNAc antennae and galactose ("LacNAc"). Capping sugars, or the terminal carbohydrates added to elongated branches of complex/hybrid glycans, include GalNAc, sialic acid, fucose, and galactose. These monosaccharides may protrude from the secondary structure of repeated sugar units (poly-LacNAc) commonly present in complex/hybrid antennae. In this manner of presentation, these carbohydrates play a central role in determining the interaction of N-linked proteins with their environment(69).

O-linked glycosylation on protein substrates is a broad term for the attachment of various carbohydrate molecules to the hydroxyl group of Ser/Thr residues(50,71). For this type of glycosylation, several subdivisions can be made by categorizing the initiating sugar attached to the polypeptide. For example, the O-GalNAc type O-glycosylation, which are commonly synthesized on cell-membrane and secreted proteins, mediate effects similar to N-glycans in cell-cell and cell-ECM interaction. Another type, O-GlcNAc, may be synthesized on cytosolic and nuclear proteins and are hypothesized to regulate signaling cascades(64). Other forms of O-glycosylation include linkages beginning with as xylose, mannose and fucose. For simplicity, relevance, and functional similarities to N-glycans, all further information presented here will be based on O-GalNAc type O-glycosylation.

The synthesis of O-GalNAc glycans on proteins (also termed "mucins") begins with transfer of the sugar molecule to Ser/Thr residues on proteins transiting through the Golgi. The reaction is carried out by the polypeptide-GalNAc-transferase family of enzymes (ppGalNAcT), who require the GalNAc molecule to be attached to the high-energy donor UDP. There is no consensus sequence required for the initiation of the reaction; however steric effects from proximal amino acids may influence the access of ppGalNAcTs to specific residues(71). Once the first GalNAc is attached to the polypeptide, additional linkages may be made using GlcNAc or galactose. This results in the production of various O-GalNAc core structures, to which other sugars may be further added ("extended core"). 8 distinct core structures exist, some of which present with differential tissue distributions. The most common arrangements, cores 1-4, are defined by secondary monosaccharide linkages to the initial GalNAc molecule. Glycoproteins carrying these types of sugars tend to be widely distributed in glandular tissues(50). Other configurations are seemingly more limited in their expression and may be associated with pathological states. For example, the minimum requirement for an O-GalNAc glycans is simply the linkage of the first GalNAc to Ser/Thr (termed "Tn Antigen"). The Tn Antigen, and its sialic-acid linked form ("Sialyl-Tn Antigen"), are rarely expressed as final glycan architectures in normal tissues, yet they have been identified in several tumor types(63,72). In a manner similar to N-glycans, the O-GalNAc glycans may be extended (core structures 1-4 and 6 only) with repeating units of GlcNAc and galactose, producing poly-LacNAc. This again allows terminal sugars such as fucose, sialic acid, galactose, and GlcNAc to protrude for roles in cell-environment interaction. Limitations in the final O- glycan structures at a given tissue are ultimately determined by

the expression/activity, distribution/order, and acceptor specificities of GTs and substrate availability(71). Importantly, these factors may be modified in different biological contexts, such as pathological conditions, and in response to external stimuli such as growth factors/cytokines.

The biological roles for the O-GalNAc glycosylation often overlap with those of N-glycans. For example, these molecules may impact cell-cell and cell-matrix interaction through protein conformation effects or by acting as ligands themselves. The importance of O-glycosylation in cell adhesion is exemplified by the Lewis blood group determinants(73). These molecules are fucosylated extended core O-glycans presented on cell membrane-based proteins which may undergo further modification by the addition of sialic acid. A vital role for the sialylated epitopes has been demonstrated in selectin-mediated adhesion, where they serve as ligands for leukocyte homing and extravasation during inflammation/tissue injury(51). Since the elucidation of this function, subsequent roles for the altered expression of sialylated Lewis antigens have been identified in invasive forms of cancer. In addition to adhesion effects, these carbohydrate moieties may protect protein substrates from protease digestion, mask antigens, and affect cell receptor turnover rates. Outside of the cell, secreted O-glycoproteins induce the formation of a protective barrier from pathogenic organisms for epithelial cell populations. This layer of mucinous glycans is composed of negatively charged sugars (sialic acid) which induce the formation of a hydration layer along luminal surfaces(74). A critical role for O-glycosylation in differentiation and development has also been observed through studies using *Drosophila melanogaster*. These investigations exposed the importance of O-glycans in integrin-mediated cell adhesion, cell polarity, and tight

junction formation during organogenesis(50). As well, this work showed that efficient O-glycosylation is required for the secretion of vital ECM and basement membrane components during development. Furthermore, transgenic murine models deficient in specific core-extending GTs present with embryonic-lethal phenotypes due to ineffective cell adhesion in vasculature development(50).

### 1.3.3 Sialic Acid

N-acetylneuraminic acid (Neu5Ac or sialic acid) is a monosaccharide frequently found at the non-reducing terminus of N-glycans and may occur as an intermediate or terminal sugar in O-glycans. The molecule is a nine carbon, acidic  $\alpha$ -keto sugar which is produced by metabolic conversion of N-acetylmannosamine (ManNAc) or from glycan salvage pathways(75). Neu5Ac is activated as a substrate for glycan linkage by reaction with the nucleotide donor CMP in the nucleus. The high energy donor molecule is transported to the cytoplasm and eventually to the Golgi, where its addition to acceptor glycan substrates are carried out by members of the sialyltransferase family(76). This group of enzymes exhibits remarkable specificity in linkage conformation and acceptor substrates; they may as well be sensitive to chemical modifications occurring on the donor sugar. Neu5Ac may be linked to other sugars (or itself) through three different anomeric configurations, with each linkage the result of specific sialyltransferases(76,77). The  $\alpha$ 2-3 and  $\alpha$ 2-6 linkages are found in both N- and O-glycans, where sialic acid is added on terminal galactose or GalNAc residues. Glycans bearing  $\alpha$ 2-3 linked sialic acid may act as substrates for the addition of other sugars on interior residues or inhibit further chain elongation(73).  $\alpha$ 2-6 configurations have additionally been identified in linkages to GlcNAc (terminal and non-terminal) as well as internal GalNAc



monosaccharides(73,78). This type of bond generally hinders supplementary sugar extensions, with a notable exception being the addition of  $\alpha$ 2-8 linked sialic acid. The final connection,  $\alpha$ 2-8, is initiated on  $\alpha$ 2-3 or  $\alpha$ 2-6 linked sialic acid for the production of polymeric chains of the sugar composed of up to 200 residues. The product of this linkage, polysialic acid (PSA), is rarely found outside of the biological process of embryogenesis or in neuronal development; however, one  $\alpha$ 2-8 sialyltransferase, ST8Sia-IV, is expressed in various tissues through adulthood(73,79). Sialic acid residues can be found in any tissue where O- and N-glycans are produced, however tissue-specific expression of certain sialyltransferases may limit the types of attachments present at a given site. The general biological properties influenced through sialic acid attachments are related to the electronegative charge of this sugar at physiological pH(76). At the cellular level, the charge state of Neu5Ac impacts protein conformation, which in turn regulates protein-protein interaction, receptor-ligand affinity and protein turnover. Sialic acid may also facilitate binding of glycan epitopes to lectins or even mask antigenic sequences from recognition. In a broad sense, the chemical attributes of this sugar affect cell-environment interactions through its distribution on both cell-membrane and ECM glycoprotein populations. Altered states of sialic acid biosynthesis and degradation are as well implicated in pathological conditions. The inborn errors of metabolism, as well as neurological disorders such as Parkinson's disease, Schizophrenia, and Alzheimer's disease, are all hypothesized to be impacted by sialylation(75,76). Furthermore, a general increase cell-surface sialylation and a shift toward production of  $\alpha$ 2-6 linkages are frequently observed on transformed cells and several lines of evidence suggest a role for Neu5Ac in tumor cell progression(63,76,78,80–83).

### 1.3.4 Glycosylation Alterations in Cancer

Aberrant glycosylation on cell-membrane and secreted glycoproteins is inherently tied to the progression of carcinoma. Alterations in the abundance, structure, and sites of glycosylation contribute to the acquisition of invasive phenotypes through effects on cell-cell adhesion, cytoskeletal dynamics, and extracellular matrix properties(53,63,78,84,85). Cancer cells acquire the ability to produce a selected subset of glycans configurations, including incomplete/truncated structures, changes in N-glycan branching patterns, and the altered presentation/secretion of mucins. In addition, increased sialylation, fucosylation, and the reappearance of embryonic-associated glycan structures (oncofetal antigens) are correlated with poor prognoses in several different cancers(63,86–89).

$\beta$ 1-6 branching on the core structure of N-glycans, catalyzed by the enzyme GnT-V, produces a scaffold upon which other sugars may be added for the production complex tetra-antennary oligosaccharides. While this process is a normal part of the N-glycan biosynthetic pathway, transformed cells often exploit this mechanism to over-produce these "bulky" molecules on cell surface-associated proteins. Hypotheses on how neoplastic cells carry out this function include direct transcriptional regulation of the gene MGAT5 during carcinogenesis. This gene, which codes for the GnT-V glycosyltransferase, contains 5' promoter region sequences which may be functionally regulated by downstream effectors of common cancer-cell signaling pathways including V-Src, Her-2, and H-Ras(63,85). In addition, TGF $\beta$ 1 has been shown to increase the stability of MGAT5 transcripts in melanoma cells, although the mechanism for this effect has yet to be uncovered(90). Various models of cancer show that these large

oligosaccharide groups may be attached on proteins including integrins, cadherins, cytokine/growth factor receptors, and other classes of membrane-bound glycoproteins(52). The biological consequence for the over-production of these structures varies with the function of the underlying protein. For example, increased  $\beta$ 1-6 branching on E-cadherin in a mouse mammary tumor model resulted in decreased homotypic interaction of the cell-cell adhesion proteins(91). The molecular rationale for these observations is most likely increased turnover of E-cadherin from the cell membrane and/or decreased protein-protein interaction conferred from the steric effects on the underlying protein. Alternatively, increased branching of N-glycans attached on cytokine receptors, like EGFR and TGF $\beta$ R, has been shown to reduce endocytosis of these molecules from the cell membrane(92,93). This outcome is hypothesized to be carried out by the stabilization of these receptors at the cell surface through lattice formation via increased interaction of these molecules with galectin-3; allowing continued signaling through these pathways. Interestingly, this specific example highlights a potential positive feedback mechanism to amplify cell programs for invasive phenotypes, granted the stabilizing effect of TGF $\beta$ 1 on MGAT5 transcripts discussed previously. The  $\alpha$ 5 $\beta$ 1 integrin, a receptor for fibronectin, is a heavily N-glycosylated membrane-bound protein integral to cell-cell and cell-ECM interactions. Guo et al. elegantly demonstrated that the over-expression of GnT-V in fibrosarcoma cells resulted in increased  $\beta$ 1-6 branching of the  $\beta$ 1 subunit of this cell adhesion protein(85). The authors established that through this mechanism cells were able to decrease attachment to fibronectin and increase migration/invasion. Furthermore, this group showed that inhibition of GnT-V catalyzed  $\beta$ 1-6 linkages, by the addition of swainsonine, reversed

these phenotypes and restored attachment to fibronectin. The glycosyltransferase GnT-III is responsible for adding a  $\beta$ 1-4 linked GlcNAc moiety to the central mannose residue of core GlcNAc N-glycan structures. The addition of this "bisecting GlcNAc" prevents complex glycan formation by blocking the action of downstream core trimming enzymes. Studies using various models of cancer have identified paradoxical roles for the GnT-III catalyzed product associated with metastatic phenotypes. For example, GnT-III expression and the presence of increased bisecting GlcNAc in ovarian cancer and hepatocellular carcinoma were able to differentiate tumor from adjacent normal tissue and were associated with higher rates of metastasis in these models, respectively(94,95). However in mouse models of melanoma, the overexpression of GnT-III resulted in a decrease of  $\beta$ 1-6 branching and correlated with less-aggressive tumor phenotypes(96,97).

The production of truncated or incomplete O-glycan structures are another feature associated with cancer-induced glycosylation changes. The normal biosynthesis of O-glycans begins with the attachment of GalNAc to Ser/Thr residues on polypeptides as they progress through the secretory pathway. Further sugar extensions are typically added to produce core and extended core O-glycans structures; which may be further modified by capping sugars such as sialic acid and/or fucose. Shorter, intermediate O-glycan architectures, such as the T and Tn antigens and their sialyated forms (sT and sTn), are almost exclusively expressed in association with pathologies such as cancer(63,84). These types of structures are often ubiquitously attached on various membrane-associated protein populations and therefore described as a generalized phenomenon rather than an effect on specific proteins. The Tn antigen (GalNAc-Ser/Thr) is considered a pan-carcinoma marker, being expressed on virtually all epithelial-based tumors(63,72,98).

This particular glycan epitope may be found on both membrane-associated and extracellular (luminal) proteins during carcinogenesis and is correlated to poor outcomes in multiple cancer types. Several mechanistic theories have been formed to explain the production of this biosynthetic intermediate. Following the addition of GalNAc to Ser/Thr by ppGalNAcT, the resulting Tn antigen may be extended by either T-synthase or C3GnT for the production of core-1 or core-3 structures, respectively. T-synthase is a widely expressed enzyme that requires the chaperone protein COSMC for appropriate Golgi localization and full catalytic activity(71,99). Reports using various models of cancer show that the loss of T-synthase or C3GnT activity, as well as COSMC mutation and epigenetic silencing, provide a means for accomplishing the overproduction of Tn antigen during tumor progression(100,101). Unlike the Tn antigen, the sialylated form of this epitope (sTn) does not serve as a precursor in the construction of other O-glycan configurations(71). sTn is produced by the action of ST6GalNAc-1 within the Golgi, where this enzyme catalyzes the  $\alpha$ 2-6 linkage of Neu5Ac to the initiating GalNAc carbohydrate. sTn expression is thought to occur as a byproduct of increased Tn production through the same mechanisms described above, because both epitopes are often identified in parallel. Investigations into alternative pathways for its accumulation, such as the over-expression or increased activity of ST6GalNAc-1, demonstrate mixed results as a means to accomplish this outcome(102–104). The T antigen ("core-1") is composed of a galactose molecule attached to the GalNAc residue of the Tn antigen (Gal $\beta$ 1-3GalNAc-Ser/Thr). Fates of the T antigen in normal O-glycan processing include GlcNAc addition for core-2 production and/or the attachment of Neu5Ac for mono-/di-sialylated T antigens(71). Increased expression of the T-antigen has been detected in

various epithelial-based tumors, usually in association Tn and sTn epitopes. However, difficulty in differentiating T antigens from similar glycan attachments on lipid substrates has hindered efforts to define the prognostic significance of these molecules in cancer progression(72). The functional relevance for the production of these incomplete O-glycan structures in cancer is evidenced by their roles in immune detection and cell adhesion. Ogata et al. established that tumor cells are able to evade natural killer cell induced lysis through the production of sTn antigens(105). In addition, others have shown that these glycan structures may inhibit interaction with cytotoxic T-lymphocytes and prevent the accumulation of antigen-presenting cells in the tumor microenvironment(106,107). A potential role for these molecules affecting cell adhesion properties has been demonstrated using prostate and breast cancer cell lines. Glinsky et al. showed a critical requirement for T antigen-mediated binding to galectin-3 on endothelial cells as a potential means of promoting metastasis(108). Another study showed that pancreatic cancer cells genetically modified to over-produce Tn antigens manifest with enhanced growth, migration, and invasive capabilities relative to controls. In this particular study, a knockout model of the COSMC chaperone incited these phenotypes which were reversed upon its re-expression(100). Studies targeting the expression of sTn have identified it on adhesion-related proteins such as CD44, integrin  $\beta$ 1, Muc1, and osteopontin; where its presence is correlated with increased metastatic potential(78). However, a few conflicting reports exist regarding these substrates and this trend between different models of cancer(109,110). Clarification of the functional mechanisms and substrate specificity for sTn attachment is therefore of great interest because of potential implications in biomarker identification and therapeutic-intervention.

The increased production of the sialyl Lewis<sup>x</sup> antigen (SLe<sup>x</sup>) by tumor cells also contributes to attainment of invasive characteristics. The SLe<sup>x</sup> motif, a tetrasaccharide structure defined by  $\alpha$ 1-3 fucose linked to GlcNAc and  $\alpha$ 2-3Neu5Ac to galactose, functions as a ligand for selectin-mediated cell adhesion during inflammatory responses(111). The epitope is expressed at low-levels on N- and O-linked glycans for non-transformed epithelial cells. However a generalized increase in the abundance of these structures, typically on O-GalNAc glycans, has been noted for several carcinomas with aggressive phenotypes. Hypotheses on how cancer cells gain a selective advantage for metastasis through this alteration are based on its function as a ligand for P-, L-, and E-selectin(63,73,78,84). Specifically, SLe<sup>x</sup>/selectin interaction is thought to promote aggregate formation of tumor cells with leukocytes and platelets in the blood stream, facilitating interactions with activated endothelium and decreasing the likelihood of immune detection(63,112). The mechanistic-underpinnings of SLe<sup>x</sup> over-expression may be rooted in dysregulated signaling from the inflammatory tumor microenvironment or the incomplete synthesis of more complex structures. From a molecular perspective, alterations in sialyltransferase/fucosyltransferase expression, epigenetic modification of genes required for further processing of these structures, and possible contributions from hypoxia-regulated pathways are all implicated in cancer-associated SLe<sup>x</sup> production(6,111,113).

Finally, transformed cells may alter the topology of mucins and display a generalized increase in cell-surface sialylation during cancer progression. Mucin 1(MUC1) is an integral membrane glycoprotein bearing both O- and N-linked glycans, which extend outward into the extracellular space. The O-glycan structures of MUC1,

which greatly outnumber N-linked, are typically configured as core-2 structures that are extended and capped with fucose and sialic acid. These extended configurations mediate an essential function of MUC1 on non-transformed epithelial cells: they provide high-levels of negative charges which allow for the formation of a gel-like, protective barrier on the luminal surface of secretory tissue(69,74). Transformed cells may exploit O-glycosylation biosynthetic machinery for the preferential production of incomplete or over-sialylated structures on MUC1. For example, breast cancer cell-derived MUC1 has been identified carrying T-antigen or sialylated Lewis structures rather than extended core-2 O-glycans(114). As a result of altered glycosylation, MUC1 may be inappropriately internalized and redistributed on basal cell surfaces where its charge state may influence cell adhesion(74,115,116). Additionally, hypo-glycosylated forms of MUC1 are suspected to allow increased access of extracellular proteases to the underlying peptide structure, resulting in liberation of the N-terminus to systemic circulation. This observation, which is also noted in states of inflammation, led to the development of a breast and pancreatic biomarker utilized in monitoring cancer treatment regimens (CA15-3)(117,118). Interestingly, the shedding of N-terminal MUC1 is thought to leave a remnant C-terminal peptide protruding from the cell membrane, which some studies suggest may propagate a signaling mechanism promoting stemness(119).

### 1.3.5 Sialic Acid in Cancer

Including, but not limited to, the expression of specific O-glycan sialylated structures (sT, sTn, SLe), is the observation that aggressive tumors often present with global increases in cell-surface sialylation. Early studies targeting the properties of cancer-cells first demonstrated the increased propensity of sialic acid-specific lectins in



binding to transformed models. Subsequent investigations aimed at clarifying a role for up-regulated Neu5Ac demonstrated that induced deficiency in its expression at the cell-surface decreased invasive characteristics and returned cells to normal growth dynamics(63,76,78). Further probing into this phenomenon identified alterations in sialyltransferases and sialidases which utilize N- and O-linked glycans, as well as lipids, as substrates(81). Furthermore, an apparent tendency for tumor cell glycoconjugates to carry sialic acid in a  $\alpha$ 2-6 linkage-dependent manner was observed(78). Hypotheses formed to explain the rationale for increased sialylation highlight the potential implications of this molecule on several biological processes, including: cell adhesion, receptor-ligand interactions, masking of glycan epitopes, and retention time of molecules at the cell surface(76,78,81). Specifically, these biological roles for sialic acid are thought to be conferred from the chemical nature of the molecule; at physiological pH, Neu5Ac carries a negative charge and may therefore mediate charge repulsion and/or influence protein conformation. Mechanistic research has provided insights into cancer-pathway regulation of sialic-acid modifying enzymes which may be partially responsible for its increased abundance. For example, the RAS oncogene has been shown to induce the expression of the sialyltransferase ST6Gal-I in a cell model of colon cancer(78,120). Other efforts demonstrate the feasibility for sialidase down-regulation in the propagation of invasive phenotypes. Uemura et al. used a transgenic model of colon cancer, capable of either NEU1 over-expression or knock-down, to show that there was an inverse relationship between sialidase levels and invasive properties(121). The identification of differentially regulated enzymes may also explain the inclination of tumor cells to preferentially catalyze  $\alpha$ 2-6 sialic acid linkage conformations. The relevance of this

attachment is evident as a means for decreased binding of extracellular glycan-binding proteins, such as galectins, to cell-surface glycoproteins bearing sialic acid in this configuration. Integrin  $\beta 1$ , for example, is a specific target for the  $\alpha 2$ -6 sialic acid linkage in cancer progression(120,122). Importantly, this cell adhesion receptor connects binding to the ECM with cytoskeletal elements through various heterodimeric interactions with other integrin subunits. In addition, the integrin molecules mediate vital intercellular signaling pathways that help cells respond dynamically to alterations in binding states. Multiple studies have highlighted differential glycosylation of the integrins occurring in parallel with altered biological scenarios, including pathological states(123,124). However, only recently have these changes been tied to specific outcomes, such as the modified binding affinity for extracellular substrates by integrin heterodimers containing Neu5Ac  $\alpha 2$ -6 linked to  $\beta 1$  subunits(120,125–128). Other observed changes in sialic acid dynamics, associated with aggressive phenotypes, include the appearance of N-glycolyneuraminic acid (Neu5Gc) and polysialic acid chains. Neu5Gc is a variant form of sialic acid in which the acetyl group on C5 of Neu5Ac is hydroxylated. This sub-type of sialic acid is not synthesized by humans due to homozygous loss of the gene responsible for the conversion of Neu5Ac to Neu5Gc; however it is produced by other mammalian species(76). Early observations indicating the presence of this structure in humans during embryogenesis and carcinogenesis led to hypotheses regarding a specific role for this molecule as an oncofetal antigen. However, further investigation indicated the presence of low-levels of anti-Neu5Gc antibodies in adults without any pathology(129). Subsequent studies showed that humans are capable of incorporating Neu5Gc into endogenous glycans by cellular uptake and processing of foreign glycans. Current

theories suggest that Neu5Gc is detected at higher levels during malignancy due to dietary intake of animal products in conjunction with the altered sialylation profiles these cells commonly exhibit. As a consequence, reactivity of endogenous antibodies to increased Neu5Gc epitopes could promote chronic inflammation and thereby serve to propagate tumor progression(82). Polysialic acid is a negatively charged linear chain of  $\alpha$ 2-8 linked sialic acid residues normally restricted to expression on select proteins in neural and immune cell populations. The structure may be linked to N- or O- glycans and is thought to influence biological functions through charge repulsion due to its poly-anionic nature. For example, during brain development the addition of these large polymers on NCAM and SynCAM1 mediate neurite outgrowth and synapse formation through inhibition of homophilic interactions between adjacent cells(79). PSA synthesis is a highly regulated process, with temporal and cell-specific expression mediated through the dynamic regulation of the polysialyltransferases ST8Sia-II and ST8Sia-IV(130). Under normal circumstances, brain PSA levels are down-regulated following specific developmental cues, leaving the PTM only in areas critical to neural plasticity. Outside of the brain, an additional role for PSA in dendritic cell chemotaxis has recently been resolved(131). The presence of PSA on tumor cell NCAM has been detected in several forms of aggressive disease, including small and non-small cell carcinomas, pituitary tumors, neuroblastoma, and Wilms tumor(130,132–135). Hypotheses formed to explain why tumor cells might express PSA are primarily based on the monosaccharides ability to interrupt cell adhesive processes. Subsequent examinations into the relationship between PSA and cancer progression have revealed correlations between polysialyltransferase levels, PSA abundance, and invasive properties in lung cancer and

tumors of neuroendocrine origin(132,134). Furthermore, inhibition of polysialyltransferase activity was demonstrated as a feasible approach to decrease migratory capabilities in some tumor models(136).

### 1.3.6 Introduction to Glycoprotein Cancer Biomarkers

The mechanisms by which tumor cells manipulate glycosylation pathways for the production of specific structures are an area of intense research. Recent studies have tied transcriptional regulation and epigenetic modification of fundamental glycozymes to commonly dysregulated cancer-signaling networks; yet, much remains to be elucidated in this realm(45,120). In the past, the functional redundancy of the glycosyltransferases was a key roadblock to dissecting how specific genetic programs could be modified for altered glycosylation. However, progress is being made through the production of highly specific transgenic cancer cell models, capable of over-expression or knockdown of specific glycan-related enzymes in combination with various oncogenic drivers. Additional efforts are focused on defining the biological consequences of altered glycosylation on specific protein targets. Importantly, it should be considered that some cancer-specific glycan structures are likely a "side-effect" of carcinogenesis and may not have functional relevance. Regardless, elucidation of the mechanisms and outcomes of dysregulated glycosylation will likely have implications for the identification of potentially targetable pathways for therapeutic intervention. Integral to these functional investigations are continued efforts to catalog cancer-specific protein glycoforms. This research is significant for diagnostic/prognostic stratification and will likely feedback to provide insights into mechanistic/functional research. Several cancer biomarkers in use today are glycoproteins, including CA15-3, CA19-9, prostate-specific antigen, CEA, and

CA-125(137). As well, therapeutic strategies based on the antigenic-properties of cancer-specific glycan epitopes have resulted in the production of anti-cancer vaccines(51). In light of newer research methods and technologies, capable of detailed characterization of complex oligosaccharides and their carriers, a resurgence of interest in identifying cancer-associated glycoforms as putative biomarkers has emerged.

### 1.3.7 Current Glycoprotein Cancer Biomarkers

MUC1, as discussed in section 1.3.4, is a transmembrane glycoprotein highly decorated with core-2 O-glycans capped with sialic-acid or fucose. In non-pathological states, MUC1 carries out an essential function in glandular tissues by facilitating the production of a mucinous layer for protection and lubrication of underlying epithelial cells. Observations that MUC1 was alternatively glycosylated and could be detected in circulation during inflammation/cancer initiated investigations into its clinical utility as a biomarker. As a result, assays were developed for the quantification of the shed portion of extracellular MUC1, which was termed Carbohydrate Antigen 15-3 (CA15-3)(118). The relevance of CA15-3 as a prognostic indicator is exemplified by the increased circulating levels of the molecule typically associated with the progression of various adenocarcinomas(138–141). For example, following surgical or chemotherapeutic intervention of breast cancer, rising levels of CA15-3 may indicate the presence of otherwise undetected micro-metastases(118,142). Unfortunately in this particular case, earlier detection of disease recurrence does not translate into delayed time to progression or improvements in quality of life for these patients. Therefore the recommendations for use of CA15-3 in breast cancer prognosis is limited to monitoring therapy in advanced disease states as a supplemental indicator for loss of response(118,142).

The identification of specific glycan epitopes have also been utilized in development of biomarker assays. For example, the detection of glycolipids or glycoproteins carrying sialylated-Lewis<sup>a</sup> structures (SLe<sup>a</sup>) may be of prognostic value in advanced gastric, colorectal, and pancreatic adenocarcinomas(143–145). SLe<sup>a</sup> is a structural isomer of the SLe<sup>x</sup> pentasaccharide whose production may be favored during carcinogenesis due to inefficient production of extended core O-glycans(111). The SLe<sup>a</sup> epitope, also termed CA19-9, was first evaluated for use as a biomarker in colon cancer following the production of cancer-specific monoclonal antibodies in the 1970's(146). Since that time, clinical utility for CA19-9 has perhaps been most well defined for pancreatic cancer; where it may be used in diagnostic and/or prognostic criteria, evaluation of therapy response, or in predicting recurrence. In reference to the diagnosis of pancreatic cancer, a comprehensive review by Ballehaninna and Chamberlain indicates values for the sensitivity and specificity of CA19-9 as 79-81% and 82-90% (respectively) in symptomatic patients(144). Low positive-predictive values generated in these studies were alleviated by incorporating imaging studies for targeted evaluation of patients presenting with pancreatic masses(147). In addition, pre-operative CA19-9 levels have been shown to correlate with pathological grade which may be useful for predicting resectable disease(148,149). Furthermore, several studies have evaluated CA19-9 use in pre- and post-operative prognostic stratification. Although multiple experimental designs were employed and variations in CA19-9 values became evident, these studies overwhelmingly support the use of this biomarker as an independent predictive value for survival(150–152). Despite the applicability of this biomarker in several clinical contexts, caution for its use is urged due to a number of limitations. Low sensitivity/specificity and

the potential for false positives preclude the use of CA19-9 as a stand-alone parameter for the diagnosis of pancreatic cancer in non-symptomatic patients. Several non-cancerous conditions, including chronic pancreatitis, gall bladder-related disease, and diverticulitis, may result in elevated levels of CA19-9(147,153). In addition, approximately 5-10% of the population are incapable of producing SLe<sup>a</sup> epitopes due to genetic variation affecting the production of an enzyme critical for its synthesis(73,144). Therefore, there is a wide range of applications for CA19-9 testing in pancreatic cancer, however the potential for false positives/negatives limit the information it provides and require guarded use in a clinical setting.

CA125, or Mucin16, is the most thoroughly explored biomarker for ovarian cancer. The molecule was first identified in the 1980's from a highly aggressive model of ovarian cancer using a monoclonal antibody(154). Subsequent investigations into the binding specificity of the Ab revealed targeting against the tandem repeat core of Muc16; in an area of heavy O-glycosylation similar to MUC1. Further development of antibodies to this peptide region resulted in the production of commercial assays for the quantification of CA125, which in turn facilitated research into its utility as a serum biomarker. Since that time, CA125 has been extensively studied in the contexts of serous-tissue based cancers, including lung and gastrointestinal; however current applications are most well defined for ovarian cancer. For this disease, examinations of CA125 levels in screening/early diagnosis indicate a positive correlation between biomarker levels and advanced disease state(155). However, as with CA19-9, clinicians are cautioned against using CA125 as a single entity for diagnosis in non-symptomatic patients due to low sensitivity/specificity(142). Efforts introduced to circumvent this limitation include

combinatorial approaches which integrate other biomarkers and imaging approaches, as well as longitudinal surveillance of CA125 levels(142,155,156). Another utilization of CA125 is in the stratification of post-menopausal women presenting with pelvic masses. For this patient population, increasing levels of the biomarker, in conjunction with other clinical criteria, may indicate the presence of malignant disease(142). However, the potential for false positives due to benign fluctuations of CA125 during ovulation/menstruation prevents its application to the pre-menopausal population(157,158). In addition, non-cancerous pathological conditions such as liver disease, inflammatory conditions, or even recent surgery may affect CA125 abundance in circulation. Relevance to monitoring therapeutic response and detection of recurrent disease have also been investigated and are currently accepted in clinical practice(142,157). Serial measurements of CA125 before, during, and after chemotherapy may be used in treatment decisions as a factor indicating the efficacy of response. For example, a decrease in 50% or more of pre-treatment levels for 28 days or greater is considered a positive indicator for therapeutic efficacy(159,160). Additionally, elevated levels of CA125 following treatment may predict relapse and help guide appropriate intervention. Finally, prognostic significance of CA125 in ovarian cancer, as an independent indicator, has demonstrated utility in predicting outcomes following surgical/therapeutic intervention. In this context, both the rate of change of CA125 levels and the amount of time the changes are observed for appear to correlate with trends in overall survival(142,157).

Alpha-Fetoprotein (AFP), a 70kDa N-linked glycoprotein, has been extensively researched as a biomarker for hepatocellular carcinoma (HCC). In normal physiological



contexts AFP is synthesized by the fetal liver and is detectable in maternal circulation during pregnancy. Following birth, AFP levels drop dramatically in both the infant and the mother and remain within a precise range in the absence of pathological circumstances. The biological functions of the molecule, while still under investigation, are hypothesized to be similar to that of albumin for fetal circulation(161). Observations that increased AFP levels are associated with hepatic pathologies in adults were first noted more than 50 years ago, prompting investigations into the clinical utility for monitoring. Initial examinations demonstrated promising results as elevated AFP levels tend to correlate well with larger liver tumors (>2cm). However upon continued evaluation, it was revealed that fluctuations in serum AFP were also occurring due to benign liver disorders(hepatitis, cirrhosis) and a low sensitivity for early-stage tumors (<2cm) also became evident. Therefore the quantification of AFP in serum for HCC surveillance was hindered by the potential for false positives due to chronic liver disease and false negatives due to inconsistent AFP elevation with small tumors(161,162). In the 1990's, glycoproteomic evaluations of AFP's single N-linked glycan, using lectin affinity methods, discerned differential binding of the molecule between alternative disease states(163). This data supported the hypothesis that differential glycan structures were preferentially attached to AFP in various pathological contexts. Ensuing investigations focused on quantification of a “cancer-specific” AFP glycoform (AFP-L3), in combination with total AFP and other biomarkers, resulting in modest increases in sensitivity for detecting early HCC(161). Current diagnostic applications of AFP utilize the biomarker as an adjunct to imaging-based methods for surveillance of patient populations at risk for HCC development. Additionally, AFP may be utilized in

monitoring in post-surgical/therapeutic contexts as well as an indicator for recurrence of undetected disease(161,164).

The advantage of monitoring glycoproteins as disease biomarkers is rooted in the nature of these molecules to dynamically respond to changes in physiological states. In addition, their inherent biosynthesis along the secretory pathway provides a convenient route for collection (e.g. serum/secreted fluids). Current biomarker assays primarily focus on assessing quantitative changes in the underlying substrate rather than directly accounting for alterations in the glycans they carry. While this approach has demonstrated value, it often fails to achieve levels of sensitivity and specificity required for confident clinical use(89,137). To overcome these limitations, efforts have been set forth to re-focus biomarker discovery research towards the elucidation of site-specific and structural glycan changes that accompany pathogenesis. A recent increase in analytical technologies available for glycoproteomic analyses, now capable of detailed characterization of the underlying peptide, glycosylation sites, and glycan compositions, has facilitated a renewed vigor in this type of approach. For example, newer mass spectrometry (MS) platforms and data-acquisition strategies provide a highly-sensitive method for global and protein-specific characterization of disease associated glycosylation changes. These innovative MS-based analyses provide a means for high-throughput, confident identification of peptides, glycans, and/or glycopeptides from samples stemming from minute amounts of complex starting materials. Together with advances in separation chemistry, enrichment/affinity/labeling techniques, and the development of bioinformatic glycan databases, detailed characterization of glycoforms/glycosites is underway.

## 1.4 Mass Spectrometry Methods for Glycoproteomics

### 1.4.1 Introduction to MS-based Glycoproteomics

Mass spectrometry based proteomic and glycoproteomic evaluations are integral to discovery phase biomarker research for several reasons. First, this approach allows for the characterization of either intact proteins or peptides, enriched or in complex mixtures, from a host of starting materials. A variety of MS instruments are available, each suited to specific applications, which may be combined with up-front techniques for optimization of experimental parameters. Modern platforms are capable of providing comprehensive qualitative and quantitative data as well as information on post-translational modifications and protein-protein interactions(165). In addition, decreased turn-around-time and moderate up-front sample preparation make this methodology well-suited for discovery-phase efforts. Previous approaches to glycoproteomics, such as 2D gels or lectin microarrays, failed to achieve the depth of coverage or sensitivity inherent to current MS systems(165). Furthermore, older MS technologies lacked the ionization efficiency and mass accuracy required to make high confidence identifications of glycan species. Newer, hybrid-type mass spectrometers, like the OrbiTrap, are capable of providing detailed glycan structural (anomeric linkages, branching patterns) and site specific information, as well as evaluating the underlying peptide backbone(166–168). In addition, modern ionization and fragmentation strategies out-perform previous methods in producing glycan fragment ions amenable to high-resolution characterization. Further supporting the use of MS for glycoproteomic applications are an abundance of options for enrichment, chemical derivatization, and chromatographic separations prior to analysis. The implementation of these techniques before MS acquisition increases the

likelihood for successful analysis by reducing sample complexity and/or modifying glycan structures so that they can be more efficiently analyzed. Several variations in MS workflows have been described, often involving different combinations of sample preparation, MS acquisition (ionization techniques, fragmentation strategies, various mass analyzers), and data processing(166,169–175). Each unique configuration has distinct advantages and disadvantages for glycoproteomic analysis and therefore should be carefully considered during experimental design.

#### 1.4.2 Enrichment Strategies for Glycoproteomics

One of the greatest challenges in MS-based glycoproteomics is derived from the fact that proteins are not equally distributed in biological fluids. This observation, known as the “dynamic range problem”, describes how a limited number of proteins dominate the proteome of secreted matrices such as serum. This may result in the inability to effectively characterize low-abundance proteins, such as those carrying PTMs, unless pre-fractionation of the starting material is carried out(176,177). Several strategies exist for the enrichment of glycoconjugates prior to MS analysis, typically targeting the chemical properties of the carbohydrate moieties as a means to selective isolate these molecules. These methods have been optimized for both bottom up (glycopeptides) or top-down (glycoproteins) approaches, allowing increased flexibility in experimental design. For example, lectin affinity, hydrazide chemistry, titanium dioxide (TiO<sub>2</sub>) and hydrophilic interaction liquid chromatography (HILIC) are all feasible means to accomplish this end.

Lectin affinity enrichment is based on the innate ability of these plant and animal-based molecules to recognize and selectively bind to particular glycan structures(89,178).

To date, over 150 lectins have been characterized with results demonstrating both broad and narrow-spectrum specificity for different glycoforms(178). For example, concanavalin A (con A) targets oligomannose structures, inferring a broad specificity of this lectin for N-glycans. Others, like *Sambucus nigra* (SNA) and L-phytohemagglutinin (L-PHA) are extremely specific in their preference for epitopes like sialic acid  $\alpha$ 2-6 linked to terminal galactose and  $\beta$ 1,6 branched N-glycans, respectively(176). Lectins may also be combined for use in one preparation to increase the robustness of initial screening efforts. Alternatively, sequential lectin affinity enrichments may be of value if a particular glycan species is of interest. Important considerations for this strategy include the experimental conditions (pH, metal co-factors) and appropriate elution strategies (competing sugars,  $\downarrow$ pH, glycosidases) for optimal results.

Hydrazide chemistry is another means to accomplish selective isolation of glycosylated substrates from complex starting materials. This procedure, often carried out in a solid-phase extraction format (SPE), relies on the chemical conversion of monosaccharide *cis*-diol functionalities to aldehydes, which then react with hydrazide groups on an immobilized support(14). Following the covalent linkage glycans to the hydrazide resin, glycosidase digestion is used to release glycan-carrier substrates for MS analysis. Although hydrazide approaches provide higher selectivity for isolating total glycan fractions relative to lectin affinity, there are a few inherent drawbacks to this method. Perhaps most obvious is the lack of information that downstream analysis can provide on glycan structure. Typical protocols use enzymatic digestion with Peptide-N-glycosidase-F (PNGase-F) to liberate the proteins/peptides, leaving the glycans bound to the immobilized support. PNGase-F treatment, which cleaves N-linked glycan structures

at the initial GlcNAc linkage to asparagine (except where  $\alpha$ 1,3 fucose is linked), results in a detectable deamidation reaction at former sites of N-linked glycosylation (179). Therefore this approach has utility for identifying underlying carrier peptides and N-linked glycosites. However, the lack of an equivalent glycosidase for releasing O-linked glycans and the irreversible chemical modification of oligosaccharide structure during the reaction limit the resourcefulness of this approach. Several reports have described the use of hydrazide chemistry in combination with other strategies, such as depletion of high abundance proteins, multiple protease digestions, and isotopic labeling methods, to successfully increase the range of information derived from single experiments(180–183).

Hydrophilic interaction chromatography (HILIC) exploits the polar nature of monosaccharide moieties for isolation of glycopeptides. This technique utilizes zwitterionic-based stationary phases which facilitate hydrogen bonding and electrostatic interactions with glycan moieties in the presence of organic solvents(184). Importantly, HILIC is amenable to downstream glycan structural analysis because no chemical alterations to the sugar structures are necessary for binding/elution(185). Subsequently this method may be used for glycomic, intact glycopeptide, and formerly-glycosylated peptide evaluations. As a drawback, a bias towards N-linked glycans may occur because of the nature of the molecules to be longer and more complex than O-glycans. Additionally, non-specific binding of peptides carrying hydrophilic amino acids, other polar biological molecules, or the presence of salt in sample preparations may reduce the efficiency of enrichment(185).

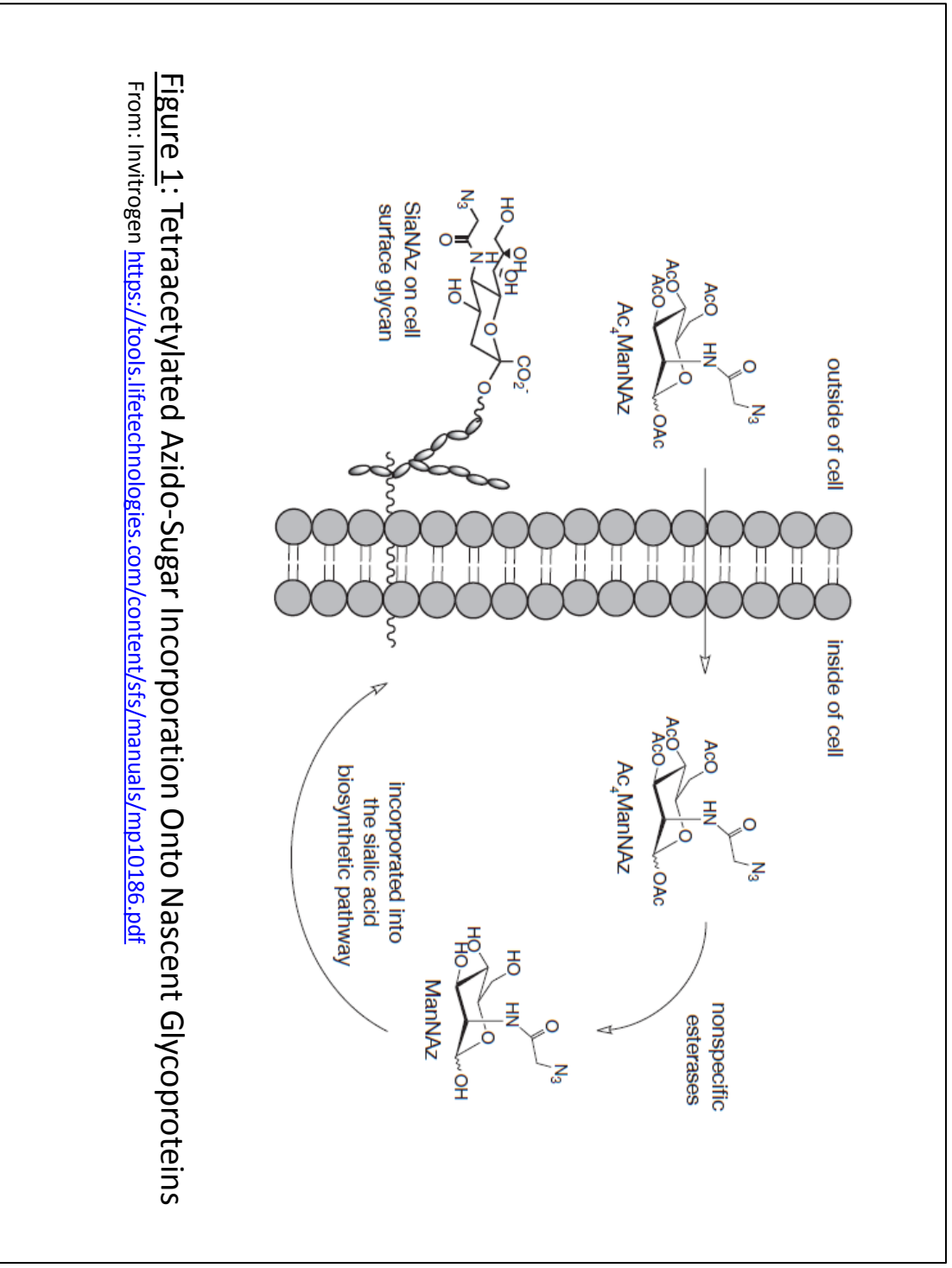
The use porous graphitized carbon (PGC) for selective enrichment of glycosylated species also has demonstrated relevance for MS applications. This method is based on the interaction of hydrophilic glycopeptides with a crystalline carbon structure through adsorption and may be used for glycomic or intact glycopeptide analyses(186). Benefits of this technique include the ability to separate isomeric glycan structures, durability, and tolerance to mobile phase pH range fluctuations(168,186). Another enrichment approach targets the chemical nature of sialic acid-containing glycans for binding and isolation of carrier substrates. Titanium dioxide ( $\text{TiO}_2$ ) is an affinity strategy originally described for phosphoproteomic applications because of its ability to bind molecules with the characteristics of a Lewis base(187). The overall negative charge carried by sialic acid-containing glycopeptides facilitates multi-point interaction with  $\text{TiO}_2$ . Therefore the use of low pH mobile phases and pre-treatment of samples with broad-specificity phosphatases are important experimental considerations. Bound substrates are eluted with increasing concentrations alkaline solutions and may be directly interfaced for MS-analysis.  $\text{TiO}_2$  chromatography is tolerant to the presence of salt and detergents in sample preparations and has been successfully combined with other enrichment methods for increased coverage of sialic-acid containing glycopeptides(188,189).

### 1.4.3 Glycan Metabolic Labeling

Glycan metabolic labeling strategies represent an alternative means to interrogate the glycoproteome for in vitro studies(190). These methods are typically based on the stable conjugation that occurs between azide and alkyne functional groups in the presence of copper(I) (“Click Chemistry”). For this type of experiment, cells are incubated with analog sugars bearing either the alkyne or azide functionality. Because these chemical

groups are small and biologically-inert, the modified sugars are incorporated onto nascent glycoproteins by normal biosynthetic pathways (**Figure 1**). Following labeling, selective enrichment/detection of modified glycoconjugates proceeds through reaction with exogenous substrates fixed with the reciprocal chemical group in the presence of copper. For example, azido-sugars may be reacted with alkynes bound to agarose beads, biotin-conjugated materials, or various fluorophores. Accordingly, this technique has applications in MS-analysis, imaging, and flow cytometry(190). Several glycan precursor analogs, including GlcNAc, GalNAc, sialic acid, and fucose, are commercially available and an increasing number of studies have taken advantage of this technology to detect alterations in both N- and O- glycosylation(191–193). This method is of particular utility in studies of secreted glycoproteins, making it highly suitable for efforts in biomarker discovery. For example, our group used the sialic acid analog (ManNAcAz) to successfully capture and quantify low-abundance glycoproteins in prostate cancer stromal cell secretomes. Remarkably, our approach maintained efficacy in MS analysis without the complete removal of serum protein supplements from the culture media(194). Similarly, Slade et al. profiled secreted glycoproteins using an azido-GalNAc analog for the detection of mucin type O-glycans stemming from CHO cell cultures(192). For MS-based applications, metabolic labeling and affinity enrichment precludes detailed glycan structural elucidation because of the irreversible nature of the Click reaction. Therefore a typical proteomic workflow relies on proteolytic release (trypsin) of non-glycosylated peptides from bound glycoproteins followed by MS identification /quantification of captured species.





**Figure 1: Tetracetylated Azido-Sugar Incorporation Onto Nascent Glycoproteins**

From: Invitrogen <https://tools.lifetechnologies.com/content/sfs/manuals/mp10186.pdf>

Further utility of this strategy by might be attained through the application a secondary digest, such as with PNGase-F, to release the bound glycopeptides for determination of potential N-linked glycosites(191). Newer developments in glycan metabolic labeling design include introduction of copper-free ligation techniques, which negates cytotoxicity and therefore may allow use of this method for in vivo studies(195).

#### 1.4.4 Strategies for MS-based Glycoproteomics

The abundance of options to selectively isolate/separate glycosylated proteins/peptides has greatly facilitated progress in MS-based biomarker endeavors. These up-front preparations may be used as stand-alone entities or in combination, offline or directly-interfaced to MS instrumentation, and approaches may be modified to suit the experimental goal. Ultimately, these techniques reduce the complexity of biological samples and thereby allow for more focused MS evaluations of glycan/peptide structures. In parallel to these advances, are improvements to MS instrumentation and novel data acquisition strategies. Importantly, these breakthroughs overcome previous analytical limitations which hindered detailed, high-throughput characterization of intact glycopeptides.

Current approaches to MS-based global glycoproteomics generally follow the scheme of: glycoprotein enrichment → enzymatic digestion of glycoproteins → chromatographic separation → MS analysis → bioinformatics. Depending on the available instruments and the experimental goal, one of two sample preparations is typically adopted: 1. Enzymatic/Chemical release of glycans from carrier peptides and separate MS acquisitions for each species (glycans and formerly glycosylated peptides) or 2. Intact glycopeptide analysis. Previous approaches to MS-based glycoproteomics

have favored the former because of difficulties in ionizing hydrophilic intact glycopeptides and inefficient detection of low m/z glycan fragment ions(196). This separate analysis approach has resulted in the development of glycomic (i.e. glycan only) MS strategies for elucidation of monosaccharide compositions and linkage conformations. As well, the formerly glycosylated peptides may be analyzed by traditional tandem mass spectrometry techniques (MS/MS) to determine amino acid sequence and localization of N-linked glycan sites. One of the major failures of the separate analysis design results from the lack of an equivalent enzyme to PNGase-F for the release O-glycans. Chemically-induced techniques, including  $\beta$ -elimination, trifluoromethansulfonic acid, and hydrazinolysis have been employed for this purpose, however low-specificity and irreversible modification to glycan/peptide structures impair characterization(171,178,197,198). In addition, chemical derivatization is almost always a necessity for glycomic analyses, adding additional time to sample preparation. The two major MS ionization strategies for glycomics, matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), both benefit from derivatization methods such as permethylation. This process decreases the hydrophilicity of glycans entering the mass spectrometer, resulting in more efficient ionization and decreasing the chances for loss of labile substituents, such as sialic acid, from the oligosaccharides(175). At the same time, formerly-glycosylated peptide analysis suffers from the potential for false identification of N-linked glycan sites. The release of N-glycans by PNGase-F results in a deamidation reaction on Asn residues carrying the oligosaccharides, producing a mass shift of .9840Da. Although this change is easily detectable by MS, the potential for random deamidations, from sample preparation procedures, limits the utility this method

for glycosite assignment(199,200). Intact glycopeptide analysis is employed to simultaneously acquire information on glycan site/structure and peptide identification. Challenges to this type of MS approach owe principally to the differing chemical properties between the peptide and the attached glycan(s). This may result in preferential data acquisition of only one of the species or complicated mass spectra containing information derived from both types of molecules. In addition, complex mixtures of glycopeptides and non-glycosylated peptides may result in the suppression of signal from the analytes of interest. Subsequently, several of MS strategies have been instituted for intact glycopeptides, including combinations of MALDI or ESI-based ionizations, single-stage or tandem MS ion fragmentations, and data acquisition with various mass analyzers(178,196). As well, bioinformatic platforms have evolved to keep pace with this type of experiment. For example, the development of glycomic and glycoproteomic data analysis softwares and curated databases containing information on N- and O- linked structures identified from various MS methods(201,202).

Of the available approaches for intact glycopeptide studies, tandem MS platforms are of particular utility because they allow for repeated cycles of precursor ion selection and fragmentation(178,203). This translates into efficient coverage of ion species representing the diversity of glycoforms present in biological samples. A general scheme for Tandem MS analysis begins with the ionization of molecules (“precursor ions”), followed by selection and fragmentation of precursor ions of interest (now called “product ions”) and finally the detection of mass to charge ratios ( $m/z$ ). The resulting  $m/z$  can then be searched against databases of known values to identify various molecular species with high confidence (amino acid sequences, sugars, lipids, etc.). Importantly, the

application of various fragmentation techniques for intact glycopeptides is used to gain information on glycan structures or site localization/peptide sequences. For example, collision-induced dissociation (CID) is a relatively low-energy approach which tends to cleave labile glycosidic bonds. The production of glycan-based spectra, with low or no fragmentation of the underlying peptide structures, makes CID suitable for glycan structural information but not for glycosite or peptide identification(173,203). Alternative CID-based methods utilize the detection of glycan-specific product (oxonium) ions for the selective isolation and secondary evaluation of precursor molecules from which they were derived(167,173,203). For example, oxonium ions produced from a first round of CID may be used to trigger a second fragmentation to acquire complementary information on glycosylation site and peptide sequence. This type of approach requires instrument capabilities to detect low  $m/z$  diagnostic ions and therefore has previously been limited to triple quadrupole and quadrupole-time-of-flight (Q-TOF) type systems; which are not ideal for discovery phase efforts. It should also be noted that the normalized collision energy used for CID approaches will influence the preferential dissociation of glycan and/or peptide structures and may therefore be modified to suit the experimental goal(173,203). Electron transfer dissociation (ETD) and electron capture dissociation (ECD) are higher energy fragmentation methods which preferentially cleave peptide backbones, leaving glycan structures intact. These MS/MS modes produce ionization patterns distinct from CID, making them favorable for peptide sequencing and identifying glycosylation sites, rather than glycan structural information(178,196,203). ECD is often carried out on Fourier Transform Ion Cyclotron Resonance (FT-ICR)-type instruments while ETD may be applied in ion-trap type mass analyzers. Both techniques

operate using similar molecular mechanisms and have utility for PTM localization studies including N- and O-glycosylation and phosphorylation(196,203). More recently, higher-energy collisional dissociation (HCD) has been employed for intact glycopeptide analysis. This type of fragmentation is specific to Orbitrap-type hybrid MS/MS systems and may be used to attain information on glycosylation sites as well as some glycan structural information(167,204). Orbitrap mass spectrometers combine linear ion trap and Fourier transform mass analyzers for the production of high resolution/high mass accuracy measurements. Technological advances in these systems allow for the detection of low m/z glycan oxonium ions from HCD; which CID-based ion trap only instruments fail to achieve. On these instruments, data acquisition strategies implementing HCD combined with CID or ETD have been used to determine glycan structure, site, and peptide sequencing in a single run(167,204,205). For discovery phase efforts, hybrid instruments are preferable over triple-quadrupole instruments because they combine high trapping capacity for ions over a wide range of masses, low mass product ion detection, and high resolution accurate mass analyzers for confident identifications.

#### 1.4.5 Bioinformatics for Glycoproteomics

Critical to the analytical advances facilitating MS glycomic and glycoproteomic investigations are developments in bioinformatic databases and software for interpreting this data. Various approaches to searching acquired mass spectra of peptide ions against known/theoretical values have been developed, each having distinct advantages/disadvantages. Proteomic databases, such as the UniProt Knowledgebase, contain information regarding amino acid sequences, PTMs, functional descriptions, and taxonomy identification. Options within this database allow users to select from manually

annotated (Swiss-Prot) or computationally-analyzed (TrEMBL) records of proteins stemming from multiple species. The application of search engine programs, such as Mascot or Sequest, facilitates comparison of experimentally-observed mass spectra against databases for the identification of peptides/proteins. However these search engine programs are limited in their ability to simultaneously consider multiple peptide modifications and have failed to keep pace with the high resolution type of data commonly acquired in modern MS systems(202). Advancements in search engine software for intact glycopeptide analysis have been made through the production of programs which consider multiple fragmentation strategies and data derived from high resolution/accurate mass systems. For example, Byonic™ software allows users to examine acquired spectra considering multiple or unexpected peptide modifications, as well as having a dedicated glycopeptide search function(202). Importantly, analyses may be set to consider data derived from multiple fragmentation strategies including HCD and ETD as well as allowing users to define monosaccharide masses or apply software-included theoretical masses. In addition to the search engine innovations for MS-based glycoproteomics has been the development of glycan structural databases. The Consortium for Functional Glycomics (CFG) contains a library of N- and O-linked oligosaccharide structures identified and validated from experimental observations. This public resource provides a platform for researchers to both query and submit data related to glycosyl transferase expression, glycan structures, protein-glycan interaction from various tissue and cell-based samples(201,206,207). The CFG glycan database be utilized to identify for glycan structures based on monosaccharide content, molecular weight values, specific glycan motifs, or linear nomenclature

(<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>).

Importantly, these options provide a means for interpreting data derived from various MS approaches (ESI-MS/MS or MALDI-MS) and sample preparations (native or derivatized glycans).



## Chapter 2

### Hypothesis

The goal of this research project was to assess the application of novel glycoproteomic and glycomic workflows in experimental approaches useful for NSCLC biomarker discovery. These methods include: 1. Azido-sugar metabolic labeling, alkyne-agarose bead enrichment and LC-MS/MS to identify proteins undergoing differential sialylation, 2. Sialic acid linkage-specific esterification and high-resolution mass spectrometry for glycan structural studies, and 3. MALDI-MS imaging of N-glycans in human lung tissue to both identify glycoforms and discern their spatial distribution.

The rationale for this work is seeded in the critical need for biomarkers to predict the disease course of NSCLC. Glycosylation is one of the most common post-translational modifications and has a well-established influence on the biological properties of glycoproteins associated with the acquisition of invasive phenotypes. Several lines of research suggest that altered glycosylation accompanies NSCLC progression and metastatic dissemination. In particular, glycans carrying sialic acid moieties are thought to contribute to the invasive properties of lung tumors through interruption of cell-cell and cell-matrix contacts. Previous efforts to attain detailed characterization of sialoglycoproteins during tumor progression have been hindered by laborious sample preparations and inefficient techniques for the selective study of these molecules. Recent advances in experimental approaches and instrumental analyses are now capable of defining glycan microheterogeneity in workflows conducive to high-throughput investigations. Cumulatively, these innovations are capable of defining glycoprotein properties including quantity, sites of glycan attachment, oligosaccharide composition, monosaccharide linkage conformations, and the architecture of glycans within tissues. Importantly, all of these characteristics could be considered when

assessing the specificity of putative glycoprotein biomarkers. Therefore, we **hypothesized** that the application of new mass spectrometry-based glycoproteomic and glycomic methods can provide a detailed characterization of sialoglycoproteins in non-small cell lung cancer. This hypothesis will be tested by the following specific aims:

**Aim 1:** Evaluate the utility of azido-sugar metabolic labeling, alkyne-agarose bead enrichment, and LC-MS/MS for identifying differentially sialylated glycoproteins in a cell model of NSCLC.

This aim entails:

- A. Optimization of the azido-sugar labeling, gene induction, enrichment, on-bead trypsin digest, and LC-MS/MS workflow. These experiments will evaluate the incorporation of the analog sugars for sialic acid (ManNAcAz) and N-acetylgalactosamine (GalNAcAz), the use of cell lysate and supernatant fractions as starting materials, and provide initial data regarding differentially captured glycoproteins following the induction of TGF $\beta$ 1, Zeb, or empty vector.
- B. Double metabolic labeling of H358 cells (SILAC and ManNAcAz), alkyne-agarose bead enrichment, and LC-MS/MS. This approach provides very reliable quantitative data on captured sialoglycoproteins following TGF $\beta$ 1 or empty vector induction.
- C. SILAC quantification of non-alkyne agarose bead enriched H358 cell lysate. Data from these experiments can be used to identify protein level changes which may be contributing to the differential capture of sialoglycoproteins.

- D. A secondary digestion using PNGase-F to release the sugar-conjugated peptides attached to the previously-trypsin digested alkyne beads. LC-MS/MS analysis of released peptides provides information on the specific glycosylation sites where ManNAcAz is being incorporated. In addition, proteins not detected following trypsin digest may also be revealed by this method.
- E. Permethylated and sialic acid linkage specific esterification of cleaved N-linked carbohydrates from H358 cell lysate followed by MALDI-FT-ICR analysis of derivatized structures. This approach will identify preliminary trends in the expression of specific sialylated N-glycans in the TGF $\beta$ 1 and EV-induced cells.

**Aim 2:** Investigate NSCLC-associated N-linked glycan structures in a set of matched human lung tissues and proximal fluids using a novel sample preparation and high-resolution MS analysis workflow.

Experiments include:

- A. Proteomic evaluation of Bronchoalveolar lavage fluid by in-solution trypsin digestion, LC-MS/MS, and label-free quantification. This catalogues the protein content of BAL which will be useful as a reference for future studies targeting the specific carriers of N-linked structures identified from glycomic evaluations.
- B. Sialic acid linkage-specific esterification and MALDI-FT-ICR structural analysis of N-glycans in pooled samples of bronchoalveolar lavage fluid and saliva. These experiments provide evidence for the feasibility of performing

glycomic studies in lung cancer proximal fluids. In addition, initial trends in the expression of sialylated structures, and their anomeric configurations, will be catalogued for reference as more samples are procured.

- C. On-tissue PNGase-F digest and MALDI-FTICR mass spectrometry imaging of N-linked glycans from human lung. Concurrent histological assessment will be performed to define regions of interest in these tissues. Direct comparison of spatial distribution of glycans to defined pathological features will be useful to identify putative cancer-specific oligosaccharide structures.
- D. Evaluate trends in the expression pattern of N-glycans between the tissues and proximal fluids. This will be useful as a preliminary assessment of how the expression of glycan structures translates between tissues of origin and associated proximal fluids.

## Chapter 3

Glycoproteomic and Glycomic Approaches to

Evaluate TGF $\beta$ 1-induced effects on

Sialoglycoproteins in a Cell Culture Model of

NSCLC

### 3.1 Introduction

Glycosylation is the most common post-translational modification, estimated to occur on approximately 50% of all proteins. Glycans influence protein-folding, localization, protein-protein interaction, and alterations in the expression/structure of specific moieties have been identified in several disease states, including cancer(63). The identification of specific glycan motifs has previously demonstrated utility for prognostic/therapeutic monitoring of various malignancies. For example, current biomarkers based on this principle include CA15-3, CA19-9, CA125, and AFP-L3(118,144,155,157,161). Despite the abundance of information these molecules carry, the clinical value of these assays are hindered by low sensitivity/specificity. Consequently, newer approaches to evaluating putative glycoprotein biomarkers are emphasizing the detection and quantification of specific glycosylation sites and structures. This work is in-part being driven by advances in up-front analytical techniques, which have multiple applications for discovery phase endeavors. For instance, glycan metabolic labeling and click chemistry-based enrichment has recently emerged as an efficient method for assessing glycosylation from in vitro models(190,195). This type of approach is based on the incorporation of analog sugars, bearing small, inert chemical functional groups, onto nascent glycoproteins. Following experimental procedures, glycoproteins which have incorporated the analog sugars are covalently bound to a solid-support and digested directly for the evaluation of proteins which incorporated them. Additionally, this method may be further employed for the identification of N-linked sites of glycosylation following the release of glycopeptides by PNGase-F digestion. In parallel to this type of experimental approach are workflow

developments and technology innovations capable of discerning individual glycan structures and, in some cases, anomeric linkages of the constituent monosaccharides. By combining these approaches for a total assessment of the glycoproteome, a more comprehensive view of how pathology influences these processes might be attained.

Therefore, we set forth to define alterations in sialic-acid containing glycoprotein populations in a cell culture model of NSCLC. An azido-sugar metabolic labeling approach was instituted, allowing biorthogonal incorporation of the analogs onto nascent glycoproteins followed by the induction of TGF $\beta$ 1, Zeb, or an empty vector control (EV). To increase the breadth of the initial screen, alterations in O-linked glycosylation were profiled by including a sample set treated with an N-acetylgalactosamine analog (GalNAcAz). Glycoproteins incorporating the azido-sugar analogs were selectively pulled down by reaction with alkyne agarose beads (i.e. “Click reaction”) and digested with trypsin. Peptides were analyzed by LC-MS/MS and normalized spectral counts were generated for quantitative evaluation of proteins incorporating the analog sugars. Next, a double metabolic labeling strategy (SILAC and ManNAcAz labels) was implemented as a secondary means to quantify captured glycoproteins following the induction of TGF $\beta$ 1. This approach was also useful for assessing if protein level changes were contributing to differential capture. To this end, we applied SILAC relative protein quantification before and after alkyne-agarose bead enrichment in TGF $\beta$ 1 and EV-induced cells. Further utility of for our methods was achieved by applying a PNGase-F digest to the previously trypsin-digested beads; allowing us to investigate the localization of N-linked glycans incorporating the azido-sugar label. Finally, the N-linked glycan structures from TGF $\beta$ 1



and EV-induced cells were evaluated using two different derivatization strategies and MALDI-FT-ICR acquisition.

## 3.2 Materials and Methods

### 3.2.1 Cell Culture

H358 non-small cell lung cancer cells were kindly provided by Dr. Robert Gemmill. Cells were grown at 37°C, 5% CO<sub>2</sub> in RPMI-1640 media supplemented with 10% fetal bovine serum and 1x Penicillin/Streptomycin/Amphotericin (MP Biomedicals). Tetracycline/Doxycycline-inducible TGFβ ligand 1, Zeb, or Empty Vector (EV) genes were stably-integrated using the Flp-In T-Rex system (Invitrogen) prior to receipt. Selection was maintained by the addition of Blasticidin (Fisher Scientific) at 5µg/ml and Hygromycin (Fisher Scientific) at 100µg/ml to media throughout culturing and experiments.

### 3.2.2 Azido-Sugar Metabolic Labeling

The workflow for H358 Azido-Sugar metabolic labeling is depicted in **Figure 2**. 1 x 10<sup>6</sup> cells were plated on 10cm polystyrene tissue culture dishes (Corning) in 6ml of RPMI-1640 media with supplements for 48 hours. At the 48 hour time-point, culture media was removed and cells were gently washed with sterile 1x PBS. Fresh media and 40µM azido sugar analog, either tetraacetylated N-azidoacetyl-D-mannosamine (ManNAcAz) (Life Technologies) or tetraacetylated N-azidoacetyl-D-N-acetylgalactosamine (GalNAcAz) (Life Technologies), were added and allowed

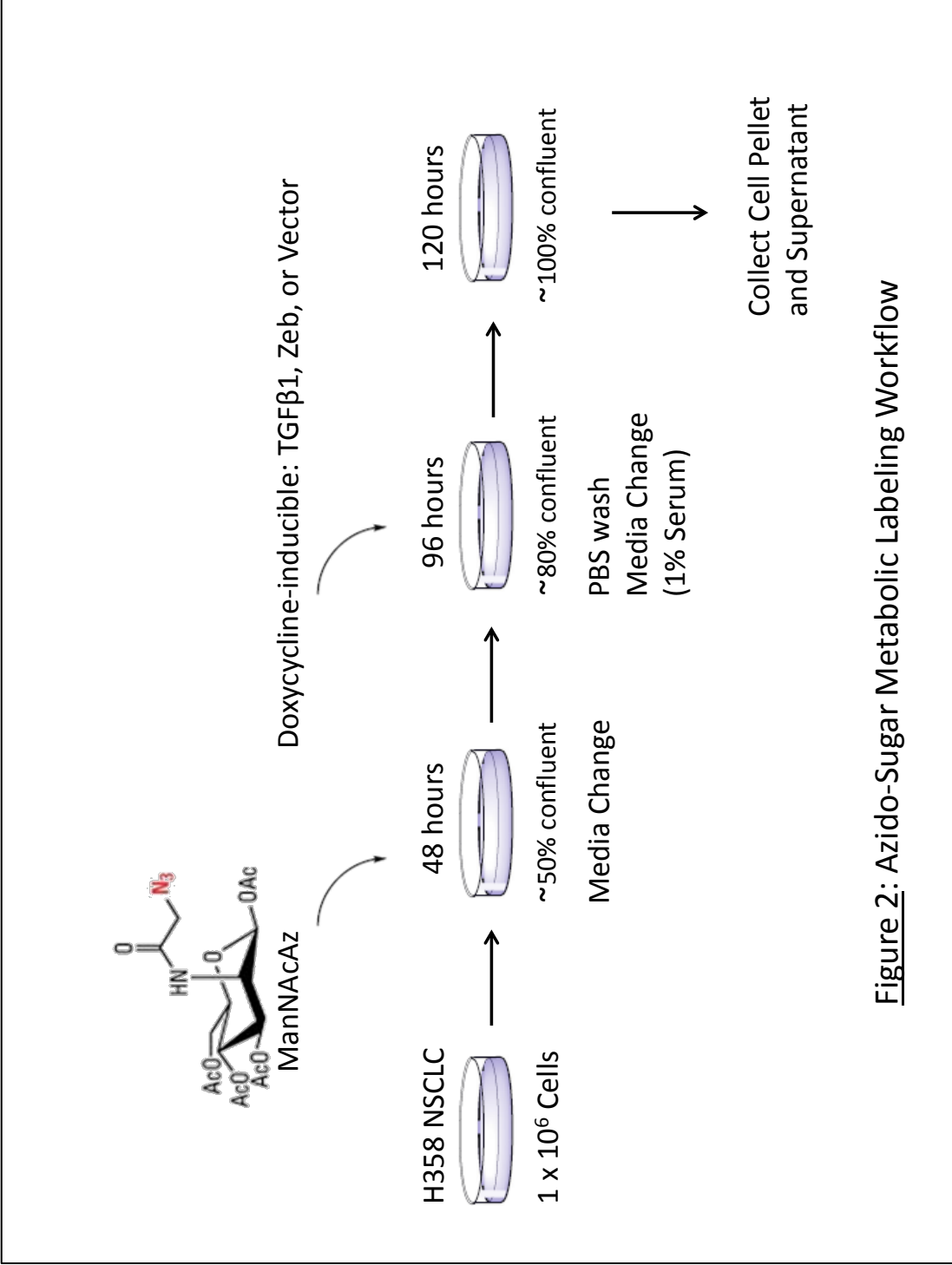


Figure 2: Azido-Sugar Metabolic Labeling Workflow

incorporation for 48 hours. At the 96 hour time-point culture media was discarded, replaced with fresh media containing 1% FBS, and 100ng/ml Doxycycline hydrochloride (MP Biomedicals) was added to the appropriate sample groups. 24 hours later, cells were loosened by non-enzymatic dissociation and the resulting culture media/intact cell mixture was centrifuged at 800rpm, 4°C, for 5 minutes in 15ml conical tubes.

Supernatant was allocated to new 15ml conical tubes and immediately stored at -80°C.

Cell pellets were washed with 5ml sterile 1xPBS, centrifuged, and the supernatant was discarded. The resulting cell pellets were immediately stored at -80°C. Three biological replicates were prepared for ManNAcAz -treated cells and two for GalNAcAz-treated.

### 3.2.3 Agarose-alkyne Bead Enrichment of Azido-Sugar bearing

#### Glycoproteins

##### 3.2.3.1 Sample Preparation

Cell pellets were processed for enrichment by combining technical replicates were combined and thawed on ice for 10 minutes in 1ml urea lysis buffer (8M Urea, 200mM Tris pH8, 4% CHAPS, 1M NaCl). Complete cell lysis was achieved through probe sonication (Fisher Scientific) at 30% power using 6x 3 second pulses while the samples remained on ice. Cell lysates were then centrifuged at 800 rpm, 4°C, 5 minutes and 800µL of the resulting supernatant was used as input for the enrichment. Conditioned cell media samples were thawed on ice and technical replicates were combined. The media samples were concentrated using Amicon Ultra 3000Da MWCO filter units (EMD Millipore) by serial centrifugations at 4°C, 2683 x g, for 45-60 minutes until the entire volume had been processed. 800µL of the resulting concentrated media was used as input for enrichment. Sample groups included in are listed in **Table 1**.

**Table 1: H358 Sample Groups for Azido-Sugar Metabolic Labeling and Enrichment, LC-MS/MS, and Label-Free Quantification**

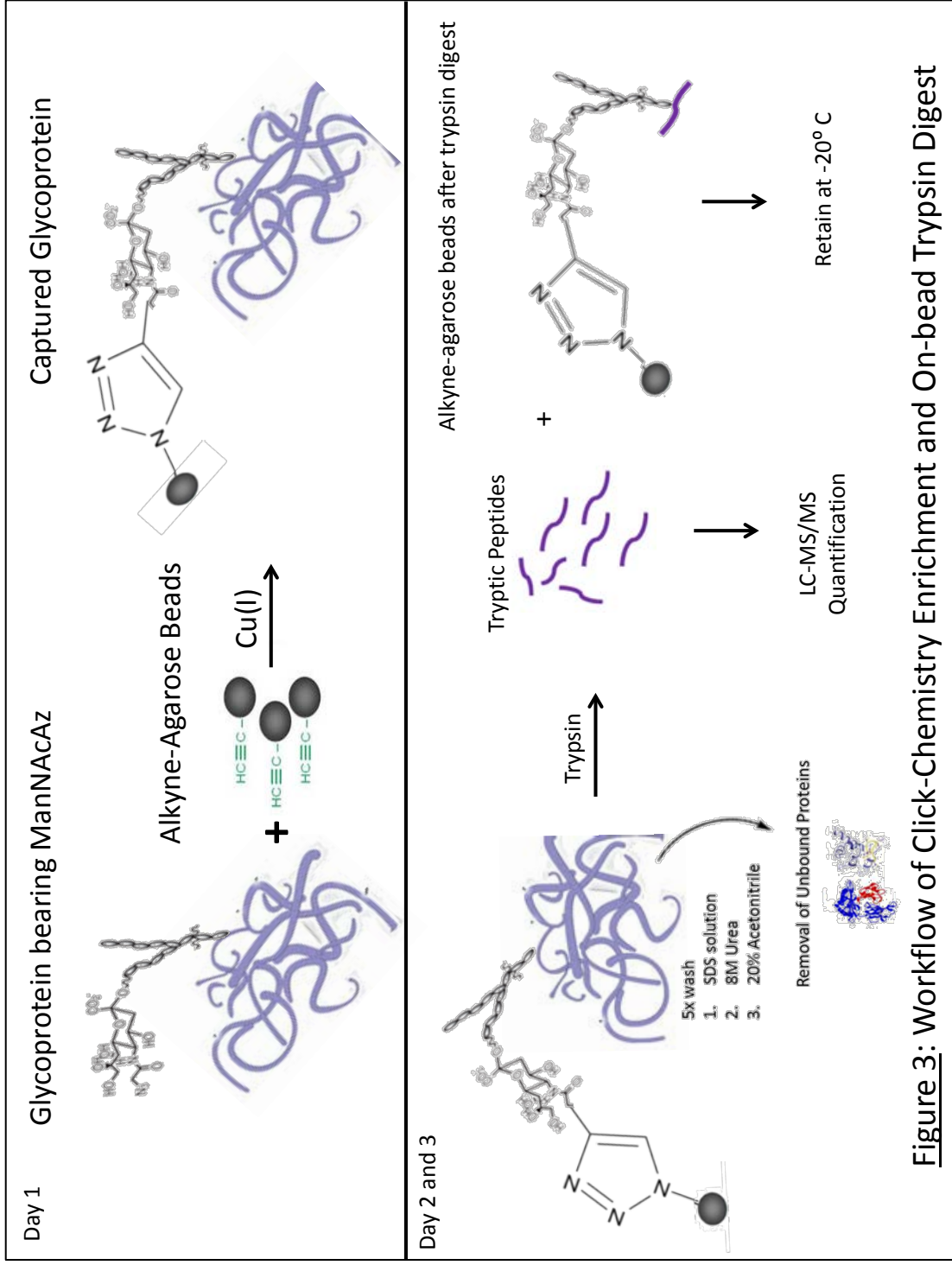
	<b>H358 Empty Vector</b>	<b>H358 TGFβ</b>	<b>H358 Zeb</b>
<b>100ng/mL Doxycycline Induction</b>	1. +40μM ManNAcAz	1. +40μM ManNAcAz	1. +40μM ManNAcAz
	2. +40μM GalNAcAz	2. +40μM GalNAcAz	2. +40μM GalNAcAz
	3. No Azido-Sugar	3. No Azido-Sugar	3. No Azido-Sugar
<b>No Doxycycline Induction</b>	4. +40μM ManNAcAz	4. +40μM ManNAcAz	4. +40μM ManNAcAz
	5. +40μM GalNAcAz	5. +40μM GalNAcAz	5. +40μM GalNAcAz
	6. No Azido-Sugar	6. No Azido-Sugar	6. No Azido-Sugar

### 3.2.3.2 Enrichment

For the enrichment of glycoproteins bearing the tetraacetylated azide sugar analogs, the Invitrogen Click-iT® Protein Enrichment Kit (#C10416) was used(208). This method encompasses a 3 day procedure in which glycoproteins bearing the azide functionality are bound to alkyne functional groups on agarose beads in the presence of a copper catalyst, producing a 1,4 disubstituted 1,2,3-triazole linkage. On the second day, agarose bead-bound samples are reduced, alkylated, and subjected to a set of stringent washes to remove unbound/non-specifically bound constituents followed by overnight digestion with trypsin. On day 3, samples are acidified, desalted, concentrated, and stored at -20°C until analysis. Reduction was performed using 10 mM Dithiothreitol (DTT) (Sigma) at 70°C and alkylation with 40 mM Iodoacetamide (IAA) (Sigma) in the dark. Washing was carried out as prescribed by the Click-iT® Enrichment protocol: 5x 2ml wash with SDS wash buffer, 5x 2ml wash with 8M Urea (Sigma), and 5x 2ml wash with 20% acetonitrile (Fisher). On-bead digestion was performed with 5µl of 0.2µg/µl proteomic grade trypsin (Sigma) for 16 hours at 37°C. Peptides were acidified by the addition of 2µL Trifluoroacetic acid (Sigma) and desalted using Waters Sep-Pak® C18 cartridges. The resulting peptides were vacuum-concentrated and stored at -20°C until LC-MS/MS analysis. Enrichment procedures are depicted in **Figure 3**

### 3.2.4 Azide-Sugar Incorporation Western Blot

H358 whole cell lysates (TGFβ and EV) derived from the ManNAcAz metabolic labeling (section 3.2.2) were lysed in RIPA buffer (10mM Tris-Cl pH 8, 1mM Ethylenediaminetetracetic acid (EDTA), 140mM Sodium Chloride, .1% Sodium dodecyl sulfate (SDS), .1% Sodium Deoxycholate, 1% Triton X-100) on ice with intermittent



**Figure 3: Workflow of Click-Chemistry Enrichment and On-bead Trypsin Digest**

probe-sonication. Protein concentration was measured by Bichinonic acid (BCA) assay from a commercially available kit (Pierce #23227). 50µg total protein was allocated for biotinylation of ManNAcAz-bearing glycoproteins by the Click-iT® Biotin glycoprotein detection kit (Invitrogen) as prescribed by the manufacturer. Biotinylated samples were speedvac concentrated and reconstituted in RIPA buffer prior to denaturation. Samples were separated by electrophoresis through 4-12% Bis-Tris SDS-PAGE (Life Technologies) and protein was transferred to a nitrocellulose membrane (LiCor Biosciences) for blotting. Following transfer, the membrane was blocked in 1:3 Licor blocking buffer diluted with 1x PBS and incubated with 1:5000 Streptavidin-IR800 (LiCor Biosciences) in blocking buffer containing .1% SDS for one hour at room temperature with rocking. The membrane was washed 5 times with 1x PBS containing .1% Tween (Fisher) and a final wash with 1xPBS without detergent. A LiCor Odyssey infrared imager (LiCor, Lincoln, NE, USA) was used to visualize the membrane. To assess consistency of protein loading, an identical SDS-PAGE gel was run and blotted with 1:5000 rabbit anti-GAPDH (Cell Signaling Technology) and 1:20,000 goat anti-rabbit IR-800 (LiCor Biosciences).

### 3.2.5 LC-MS/MS and Bioinformatics for Label-Free Quantification

Concentrated digests were resuspended in 15µl buffer A (97.8% HPLC-grade H<sub>2</sub>O, 2% acetonitrile, 0.2% formic acid), the entire contents of which was loaded onto a 25cm x .0075mm C-18 reversed phase LC column (packed in house, Waters ODS C18). Liquid chromatography (LC)-electrospray ionization (ESI) -tandem mass spectrometry (MS/MS) on a linear ion trap mass spectrometer (LTQ, ThermoFischer) coupled to a LC Packings nano-LC system was performed utilizing a 120 minute linear gradient from 5%

acetonitrile, 0.2% formic acid to 50% acetonitrile, 0.2% formic acid. Data-Dependent Analysis was selected to perform MS/MS on the 10 most intense ions in each MS spectra with a minimum ion count of 1000. Acquired data were searched using Mascot (version 2.4.1, Matrix Science, London, UK) via Proteome Discoverer (version 1.4.0.288, Thermo Scientific). The data were searched against a UniProt SwissProt *Homo sapien* database (2014\_07 release) containing both canonical and isoform entries, which was concatenated with entries from the common Repository of Adventitious Proteins (40,872 entries). Precursor mass tolerance was set to 2 Da and fragment mass tolerance to 1 Da. Enzyme specificity was trypsin, allowing for two missed cleavages. Carbamidomethyl (Cys) was specified as a fixed modification and protein N-term acetylation, deamidation (Asn, Gln), Gln→pyro-Glu (N-term Gln), and oxidation (Met) were set as variable modifications. False positive rates were determined by searching against an automated decoy database within Mascot. Protein identifications were confirmed using ProteoIQ (version 2.3.08, Premier Biosoft, Palo Alto, CA) by specifying >1 peptide and protein false discovery rate (FDR) <5%. Protein FDR was calculated within ProteoIQ using separate searches against target and reverse decoy databases. The decoy database was the target database reversed. ProteoIQ was used to generate normalized spectral counts (nSCs) for use as relative quantification of protein expression levels for comparisons between cell groups. Designation of “Glycoproteins” was done by cross-referencing identified proteins against the glycoprotein keyword, KW-0325, in the UniProt Knowledge Base. A moderated t-test (limma package, R) was used to determine significant changes, which was defined as  $p < 0.05$ . These significant proteins were uploaded to the WebGestalt (<http://bioinfo.vanderbilt.edu/webgestalt/>) in order to perform pathway enrichment



analysis. Enrichment was performed against the KEGG database with the entrez protein-coding genes as a background. A hypergeometric test was used and required a minimum of two proteins for a pathway. Significant pathways were identified if Benjamini-Hockberg (BH) adjusted  $p < 0.05$ .

### 3.2.6 Double Metabolic Labeling

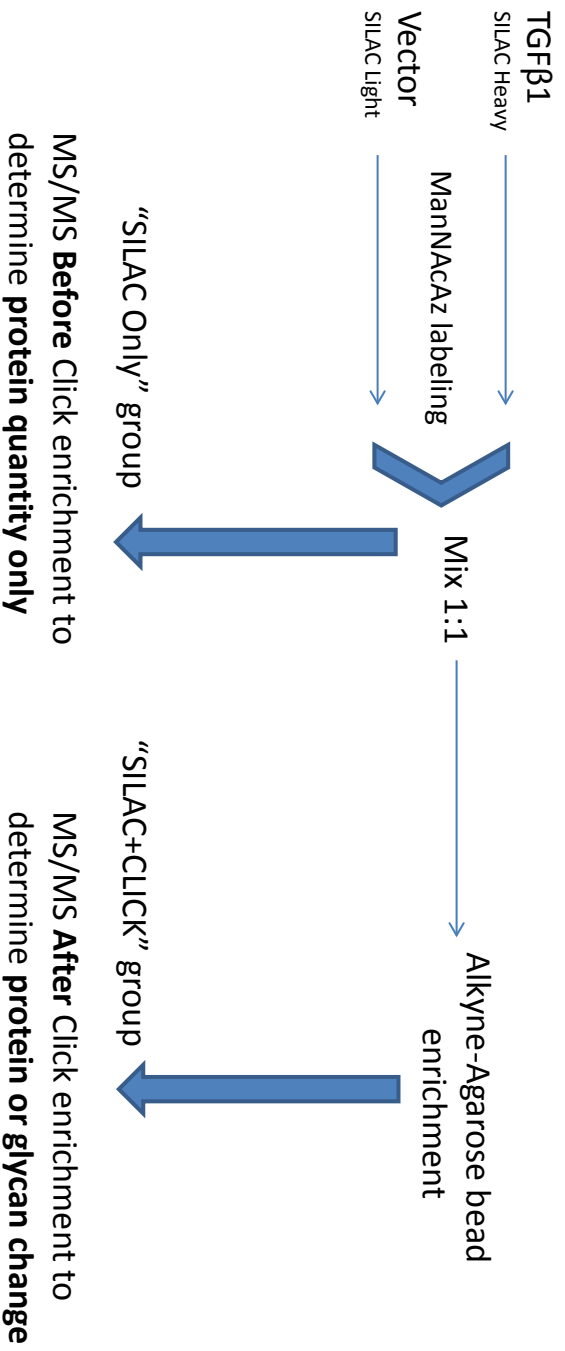
TGF $\beta$  and EV-inducible H358 bronchoalveolar adenocarcinoma cells were selected for a double-metabolic labeling approach encompassing: 1. Stable-isotope labeling of amino acids in cell culture (SILAC), followed by 2. Azido-sugar metabolic labeling using ManNAcAz. For SILAC labeling, cells were cultured using Pierce<sup>®</sup> SILAC protein quantitation kits (#89982) with <sup>13</sup>C<sub>6</sub>-L-Lysine and <sup>13</sup>C<sub>6</sub>-<sup>15</sup>N<sub>4</sub>-L-Arginine (“Heavy”) or non-isotopic L-Lysine and L-Arginine (“Light”) in T75 tissue culture flasks for 6 doublings. Incorporation of heavy amino acids was verified >95% by MS prior to beginning the azido-sugar metabolic labeling procedures. Azido-sugar metabolic labeling was carried out as described in section 3.2.2, with the following changes: 1. Cells were maintained in full (dialyzed) serum SILAC media throughout and, 2. 40ng/ml Doxycycline was used for induction of TGF $\beta$  or EV. A total of three biological replicates, including a label-swap, were prepared for these experiments (**Table 2**).

#### 3.2.6.1 Sample Processing for Double Metabolic Labeling

The rationale for our double metabolic labeling experiments is depicted in **Figure 4**. Following double metabolic labeling procedures, cells were collected and stored at -80°C as described in section 3.2.2. Pellets were processed for either: 1. SILAC relative quantification before agarose-alkyne bead enrichment (“SILAC-Only”), or 2. SILAC relative quantification after agarose-alkyne bead enrichment (“SILAC+Click”). SILAC-

**Table 2: H358 Sample Groups for Double Metabolic Labeling and Protein Quantification (Before and After Enrichment)**

	<b>H358 Empty Vector</b>	<b>H358 TGFβ</b>
<b>All samples: +40ng/mL Doxycycline</b>	1. SILAC Heavy Label +40μM ManNAcAz	1. SILAC Heavy Label +40μM ManNAcAz
	2. SILAC Light Label +40μM ManNAcAz	2. SILAC Light Label +40μM ManNAcAz



**Figure 4:** Rationale for Double Metabolic-Labeling Approach. SILAC-Incorporated TGFβ1 and EV cells undergo the 120 hour azido-sugar metabolic labeling and gene induction procedure and are then mixed at a 1:1 ratio. Two fractions are prepared for MS acquisition: 1. Without enrichment (“SILAC-Only”) and 2. With enrichment (“SILAC+Click”)

Only pellets were thawed on ice in 500 $\mu$ L RIPA buffer for 10 minutes followed by probe sonication at 30% power for 6 seconds with three repeats. Samples were centrifuged 800rpm, 4 $^{\circ}$ C, for 5 minutes and approximately 450 $\mu$ L of the resulting supernatant was saved to 1.5ml microcentrifuge tubes (Axygen) on ice. Protein quantification was achieved using the BCA assay and 100 $\mu$ g SILAC pair aliquots were prepared (50 $\mu$ g SILAC Heavy + 50 $\mu$ g SILAC Light) for in-solution trypsin digestion. Samples were reduced at 95 $^{\circ}$ C with 10mM DTT for 5 minutes followed by alkylation with 40mM IAA in the dark for 30 minutes. 2 $\mu$ g of Trypsin was added to each SILAC pair sample and digestion proceeded at 37 $^{\circ}$ C overnight. The resulting peptides were desalted using Pierce C18 spin columns (#89870), speedvac concentrated, and stored at 4 $^{\circ}$ C until LC-MS/MS. For SILAC+Click samples, normalization of input for Heavy and Light pairs was accomplished by cell counting immediately following non-enzymatic dissociation. 5 x 10<sup>6</sup> total cells (2.5 x 10<sup>6</sup> Heavy + 2.5 x 10<sup>6</sup> Light) were prepared for agarose-alkyne bead enrichment by the addition of 500 $\mu$ L Urea lysis buffer to each pellet, vigorous pipetting, and combining pairs to attain 1000 $\mu$ L total sample volume. Enrichment of ManNAcAz-bearing glycoprotein was carried out as described in section 3.2.3 and samples were stored at 4 $^{\circ}$ C until LC-MS/MS analysis.

### 3.2.7 LC-MS/MS and Bioinformatics for Double Metabolic Labeling

Concentrated digests were resuspended in 7 $\mu$ L buffer A (97.8% HPLC-grade H<sub>2</sub>O, 2% acetonitrile, 0.2% formic acid) for online 2-dimensional liquid chromatography. Samples were first loaded onto a strong cation exchange column (Biobasic, 50 x .5mm, Thermo) using a Dionex U3000 nano-LC system (Thermo Scientific). For SILAC-Only samples, 10-step SCX separation was carried out using increasing concentrations of

ammonium acetate (0, 5, 10, 20, 40, 50, 60, 75, 100, and 500mM) for 10 minutes at each concentration. For SILAC+Click samples, a 5-level SCX fractionation was utilized (0, 5, 10, 50, 500mM). Peptides were eluted at each step directly onto a C18 trap column (C18 pepMap 100, 5 $\mu$ m, 300 $\mu$ m x 5mm ID, Thermo) followed by a C18-reverse phase analytical column (75mm ID, 15cm) packed in-house with Waters YMC-ODS C18-AQ 5mm resin. Peptides were eluted from the analytical column to an OrbiTrap Elite instrument (Thermo Scientific) using a 120 minute linear gradient of 2-50% buffer B (97.8% acetonitrile, 2% HPLC-grade H<sub>2</sub>O, .2% formic acid) at a flow rate of 200nL/min via a Nanospray Flex Ion source (Thermo Scientific). Data-dependent analysis was selected to perform MS/MS on the 20 most intense ions in each MS spectra with a minimum ion count of 1000. The survey scan was performed in the Orbitrap at a resolution of 60,000. For ion trap based CID, the normalized collision energy was set to 35% and dynamic exclusion was set to a repeat count of 2, repeat duration of 30 seconds, and exclusion duration of 180 seconds. Acquired data was searched using MaxQuant quantitative proteomics software (version 1.5.1.0) against a UniProt Swissprot *Homo sapiens* database (2014\_7 release), which was concatenated with entries from the common Repository of Adventitious Proteins (40,872 entries total). Precursor mass tolerance was set to 4.5 ppm and fragment mass tolerance to 0.5 Da. Enzyme specificity was trypsin and 2 missed cleavages were allowed. Carbamidomethyl (Cys) was selected as a fixed modification and protein N-term acetylation, deamidation (Asn, Gln), pyro-Glu (N-term Gln), and oxidation (Met) were specified as variable modifications. The false discovery rate was set to 1% using a reversed target database generated in MaxQuant. Heavy SILAC labels were designated as Arg 10 and Lys 6 with a maximum of 3 labeled

amino acids per peptide. Protein quantification required a minimum ratio count of 2 peptides, including all unmodified and variably-modified peptides (protein N-term acetylation and oxidized methionines). SILAC ratios generated in MaxQuant were post-processed using Perseus statistical software (version 1.4.0.8). Reverse, contaminant, and “identified by site only” hits were filtered and removed. Replicate trimming was applied by calculating  $\pm 2$  standard deviations of the mean of the log<sub>2</sub> ratio distribution for every label-swap comparison (Heavy/Light\*Light/Heavy) and discarding proteins falling outside of that range(209,210). Ratios were inverted so that all data was reported with respect to TGF $\beta$  expression and log<sub>2</sub> transformed. A moderated t-test (limma package, R) was used to determine significant changes ( $n \geq 2$  of 3 ratios), which was defined by Benjamini-Hockberg corrected  $p$  values. KEGG analysis of significant proteins in SILAC+Click group was carried out as described in section 3.2.5.

### 3.2.8 PNGase-F Mediated Release of N-linked Glycopeptides

Agarose-alkyne beads retaining ManNAcAz-bearing glycopeptides following on-bead trypsin digestion (section 3.2.3.2) were stored at -20°C until processing. Beads were initially washed with 100 $\mu$ L RIPA buffer, vortexed, briefly centrifuged, and the supernatant was aspirated with a gel-loading tip and discarded. 100 $\mu$ L of HPLC H<sub>2</sub>O and 1 $\mu$ g of recombinant Peptide-N-glycosidase-F (See Powers et al for enzyme information (211)) were added followed by incubation for 16 hours at 37°C. Following digestion, the supernatant was collected and the beads were washed with an additional 100 $\mu$ L of HPLC H<sub>2</sub>O, which was added to the supernatant. The resulting 200 $\mu$ L aliquots were centrifuged briefly to pellet any carryover beads and the supernatant was speedvac concentrated.

### 3.2.9 LC-MS/MS and Bioinformatics for PNGase-F Mediated Release of N-linked Glycopeptides

Concentrated peptides were reconstituted in 7 $\mu$ L Buffer A and analyzed on the OrbiTrap Elite using the parameters described in section 3.2.5; with the exception that only a single dimension of chromatography was carried out (RPLC). Acquired data was searched using MaxQuant as described in section 3.2.7, with one peptide being required for protein identification. Post-processing using Perseus software was used to generate categorical headings (i.e. peptide sequence, deamidation site, protein ID) and data was exported to external data-processing software for the identification of putative N-linked glycosylation sites by manual annotation.

### 3.2.10 H358 Global N-linked Glycomics

#### 3.2.10.1 Linkage-Specific Sialic Acid Esterification

TGF $\beta$  and EV inducible H358 cells (+/- Doxycycline induction) were prepared as described in section 3.2.2, with the exception that no azido-sugar metabolic label was added. Samples were incubated at 4 $^{\circ}$ C in 200 $\mu$ L of lysis buffer (1% SDS, 5mM EDTA, 25mM Tris Base) for 1 hour. Probe sonication was performed at 30% power, 10 seconds and samples were centrifuged for 5 minutes, 800rpm, 4 $^{\circ}$ C. Supernatant was transferred to new tubes and protein quantification was carried out by BCA assay. 250 $\mu$ g of protein was allocated to a fresh tube and total sample volume was brought up to 100 $\mu$ L in water.

Protein precipitation was achieved by adding 400 $\mu$ L methanol (Fisher), 100 $\mu$ L chloroform (Fisher), 300 $\mu$ L HPLC H $_2$ O (Fisher), vortexing, and centrifugation at 14,000 x g for 2 minutes. Supernatant was discarded and 400 $\mu$ L of methanol was added to the protein pellet followed by vortexing and centrifugation at 14,000 x g for 3 minutes. The

entire solution was decanted and protein pellets were allowed to air dry for 10 minutes. 100µL of reducing solution (10mM DTT, 1% SDS) was added and dissolution of protein pellets proceeded by rigorous pipetting and vortexing. Samples were boiled for 10 minutes and cooled to room temperature. Nonidet P-40 (Boehringer) was added to a final concentration of 1%, as well as 1µg of PNGase-F. Digestion was carried out at 37°C for 16 hours. On the next day, protein was re-precipitated by the addition of 400µL methanol, 15 minute incubation at -80°C, and centrifugation at 15,000 x g for 10 minutes. At this point, the supernatant (containing released N-linked glycans) was collected and speedvac concentrated. 2µL of HPLC H<sub>2</sub>O and 40µL of Ethylating reagent ( 250mM 1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide hydrochloride (EDC), 250mM 1-Hydroxybenzotriazole hydrate (HOBt)) (Oakwood Chemical) were added, followed by incubation at 37°C for 1 hour. After incubation, 40µL of Acetonitrile was added and samples were incubated at -20°C for 20 minutes. Cotton tip HILIC solid-phase extraction of N-glycans was performed using the procedure described by Reiding et al and samples were eluted in 10µl HPLC H<sub>2</sub>O (212).

#### 3.2.10.2 Permethylation

H358 TGFβ and EV cell pellets (+/- Doxycycline induction) were probe-sonicated in 1x PBS at 30% power for 10 seconds on ice. Protein was precipitated by the addition of cold methanol (80% final concentration) and centrifugation at 20,000 x g, 4°C, for 30 minutes. Supernatant was removed and the protein pellets were reconstituted in 200µl 1x PBS. Protein was re-solubilized by vigorous pipetting and each sample was split into two 100µl aliquots: one for PNGase-F (2.7µg per sample) digest and one as a no-enzyme control. Following overnight incubation at 37°C, 400µl of cold methanol was added and



samples were centrifuged at 20,000 x g, 4°C, for 30 minutes. The supernatant, containing the glycans, was speedvac concentrated and permethylation was carried out as described by Mechref et al.(213). Following permethylation, samples were speedvac concentrated and reconstituted in 10µl of 50% methanol.

### 3.2.11 MALDI-FT-ICR Analysis and Bioinformatics of N-linked Glycans

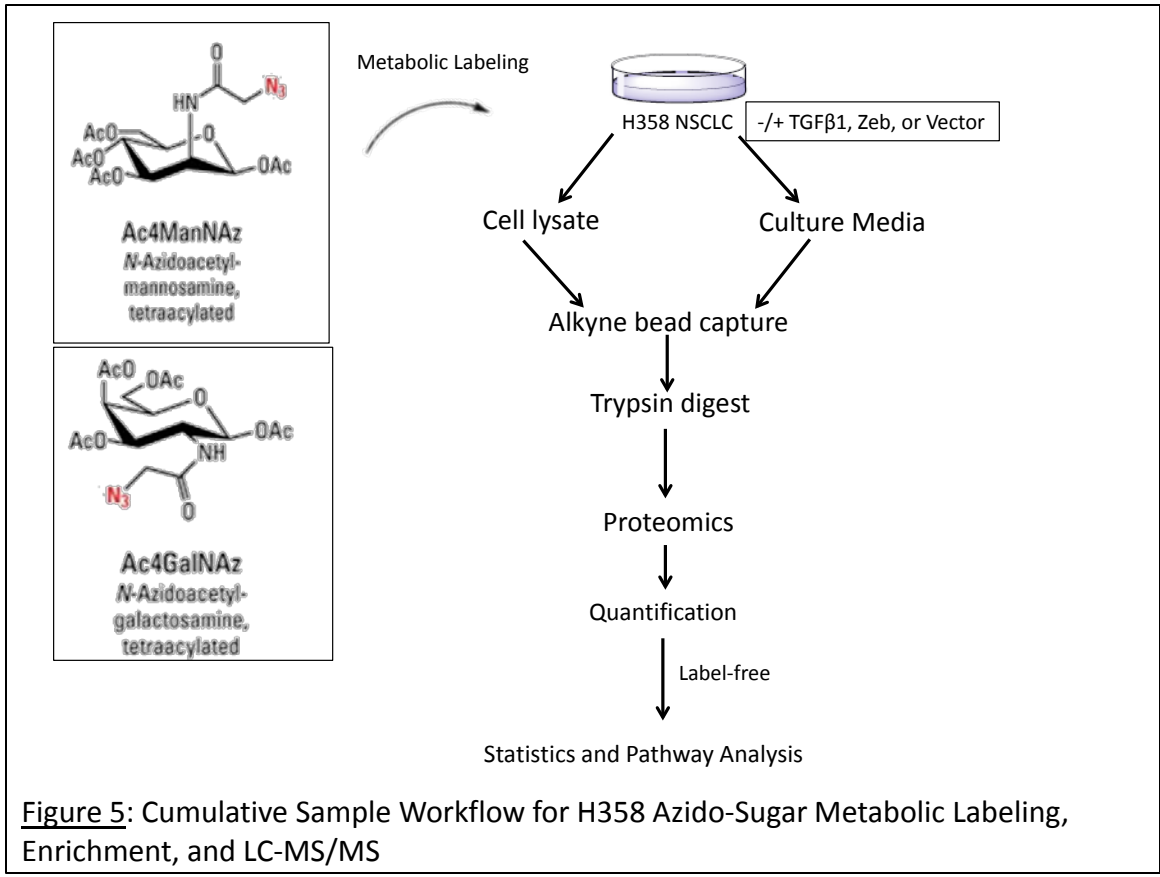
5µL of the Cotton tip HILIC eluate were spotted on an Anchorchip MALDI plate (Bruker Daltonics) with .3µL DHB matrix (5mg/ml 2,5-Dihydroxybenzoic acid (Sigma) in 50% acetonitrile, 1mM NaOH (Sigma)). After each spot dried, .2µL of ethanol (Fisher) was added for recrystallization. For Permethyated glycans, 1µl of sample and 1µl of DHB were spotted and allowed to air dry. Spectra were acquired on a Solarix dual source 7T FT-ICR mass spectrometer (Bruker Daltonics) in positive mode ( $m/z= 500-5000$ ). A SmartBeam II laser was operated at 1000 Hz in random walk mode for the acquisition of 10 average scans per spot, 75µm laser spot size. Acquired spectra were visualized using FlexImaging 4.1 (Bruker Daltonics) and exported to DataAnalysis 4.0 (Bruker Daltonics) for peak selection. Annotation of glycans was carried out using GlycoWorkbench 2.1 by searching spectra against the glycan structures database from the CFG(<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>)(214). Sialic acid linkage specificity was determined by identifying mass shifts of -18Da ( $\alpha$ 2-3) and +28Da ( $\alpha$ 2-6) relative to the non-derivatized mass of identified species as described by Reidling et al.(212).

## 3.3 Results

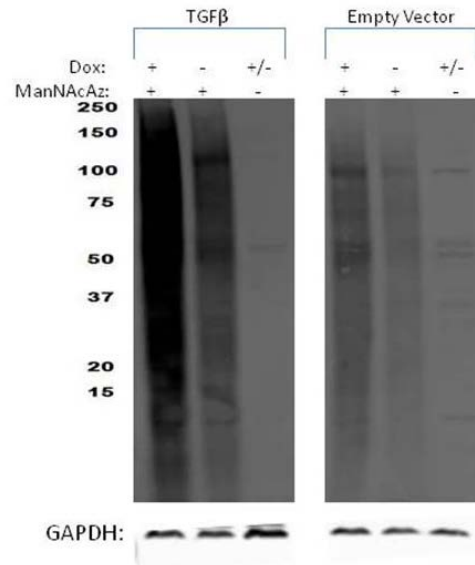
### 3.3.1 Initial Quantification of Glycoproteins in TGF $\beta$ 1, Zeb, and EV-induced NSCLC Cells

We first sought to identify how the induction of TGF $\beta$ 1, Zeb, or EV would affect glycoprotein populations bearing either sialic acid or O-GalNAc moieties in the H358 NSCLC cell line. An azido-sugar metabolic labeling strategy and alkyne-bead enrichment was utilized for the selective incorporation and pull-down of proteins potentially undergoing alternative glycosylation following induction of these genes. The workflow for these experiments can be summarized as: 1. Metabolic labeling of the TGF $\beta$ 1, Zeb, and EV-inducible cells with either tetraacetylated N-azidoacetyl-D-mannosamine (ManNAcAz) or tetraacetylated N-azidoacetyl-D-N-acetylgalactosamine (GalNAcAz) and induction of gene expression, 2. Ligation of cell-lysate derived glycoproteins bearing the azido-sugars to alkyne-agarose beads via Click chemistry and removal of non-specific proteins, 3. Trypsin digest off of beads, LC-MS/MS, and label-free quantification bound glycoproteins (**Figure 5**).

Prior to MS/MS analysis, an assessment of azido-sugar incorporation efficiency in the H358 cells was carried out. Cells were incubated in the presence of ManNAcAz for 48 hours followed by doxycycline-induction of genes for 24 hours. Cell lysate was incubated with biotin-alkyne, separated via SDS-PAGE, and blotted using streptavidin-IR dye. As illustrated in **Figure 6**, the incorporation of ManNAcAz is evidenced by the



**Figure 5:** Cumulative Sample Workflow for H358 Azido-Sugar Metabolic Labeling, Enrichment, and LC-MS/MS



**Figure 6:** H358 Cells Incorporate ManNAcAz. Cell lysate from the azido-sugar metabolic labeling procedure was incubated with biotin-alkyne. Approximately 12.5µg was separated by 4-12% Bis-Tris PAGE, transferred to nitrocellulose, and incubated with streptavidin-IR800.

presence of a broad smearing in both doxycycline-induced and non-induced TGF $\beta$  and EV H358 cells.

For MS/MS-based label-free quantification of captured glycoproteins, a total of 1091 proteins were identified from all samples, including cell lysate and supernatant fractions, TGF $\beta$ , Zeb and EV-induced cells, and labeling with both ManNAcAz and GalNAcAz (**Table 3, suppl. materials**). Of the 1091 proteins, 387 matched the UniProt knowledge base keyword “glycoprotein” (KW-0325), 24 were isoform-specific identifications, and 17 proteins were identified as both canonical and isoform-specific. From this data, individual comparisons were made between the cell-lysate derived experimental groups by applying a moderated t-test to the averaged log<sub>2</sub> fold change of normalized spectral counts (nSC). Because significance was being deemed by the averaged fold change metric, 50% of the minimum observed nSC value (MOV) was applied for any protein which was not quantified in at least one biological replicate of a given comparison, so that no value would be zero. **Table 4** summarizes the total number of proteins quantified, those reaching statistical significance ( $p < 0.05$ ), and how many were up or down regulated for all comparisons.

The induction of TGF $\beta$  following ManNAcAz-labeling resulted in the quantification of 598, 611, and 558 proteins relative to induced-EV, non-induced TGF $\beta$ , and induced-Zeb expression, respectively (**Tables 5 – 7, suppl. materials**). The corresponding numbers of proteins which were captured at significantly different levels were 10, 16, and 57, respectively. The proportion of glycoproteins identified in these experiments was 46% for both TGF $\beta$  vs. EV and induced vs. non-induced TGF $\beta$  and 47% for TGF $\beta$  vs. Zeb. KEGG analysis revealed 3 enriched pathways (BH $p < .05$ ) from

Table 4: Summary of H358 Cell Lysate Label Free Quantification Results									
Experimental Groups Compared		Total # of Proteins Quantified	# Significant ( $p < .05$ )	Relative to Group 1 Induction:					
Group 1	Group 2			Total # Up-regulated	# Up-regulated $p < .05$	Total # Down-regulated	# Down-regulated $p < .05$		
H358_TGFβ1 +Dox +MannNAcAz	H358_EV +Dox +MannNAcAz	598	10	297	4	301	6		
H358_TGFβ1 +Dox +MannNAcAz	H358_TGFβ1 No Dox +MannNAcAz	611	16	295	9	316	7		
H358_TGFβ1 +Dox +MannNAcAz	H358_Zeb +Dox +MannNAcAz	558	57	457	57	99	0		
H358_TGFβ1 +Dox +GalNAcAz	H358_EV +Dox +GalNAcAz	455	14	195	6	260	8		
H358_TGFβ1 +Dox +GalNAcAz	H358_TGFβ1 No Dox +GalNAcAz	463	9	188	3	275	6		
H358_Zeb +Dox +MannNAcAz	H358_EV +Dox +MannNAcAz	543	50	90	0	453	50		
H358_Zeb +Dox +MannNAcAz	H358_Zeb No Dox +MannNAcAz	525	51	145	1	380	50		

the proteins captured at significantly different levels in induced TGF $\beta$  vs. induced-EV, 4 pathways with induced vs. non-induced TGF $\beta$ , and 27 for TGF $\beta$  vs. Zeb expression. When TGF $\beta$  was induced in GalNAcAz-labeled cells, 455 and 463 proteins were quantified, relative to EV and non-induced TGF $\beta$  cells (**Tables 8 and 9, suppl. materials**). Of these quantified proteins, 14 and 9 were captured at significantly different levels, respectively. A total of 31% and 30% of the total quantified proteins matched the UniProt keyword search for glycoprotein and KEGG analysis identified three significantly enriched pathways related to TGF $\beta$  induction in the GalNAcAz-based experiments. Comparisons of ManNAcAz-labeled induced-Zeb versus induced-EV and non-induced Zeb cells (**Tables 10 and 11, suppl. materials**) resulted in a total of 543 and 525 proteins quantified, respectively. Statistical significance was reached for 50 and 51 of the proteins quantified in these comparisons. The proportion of total glycoproteins was 48% and 54%, and KEGG analysis identified 19 and 27 significantly enriched pathways for these experiments, respectively.

### 3.3.2 Double Metabolic Labeling

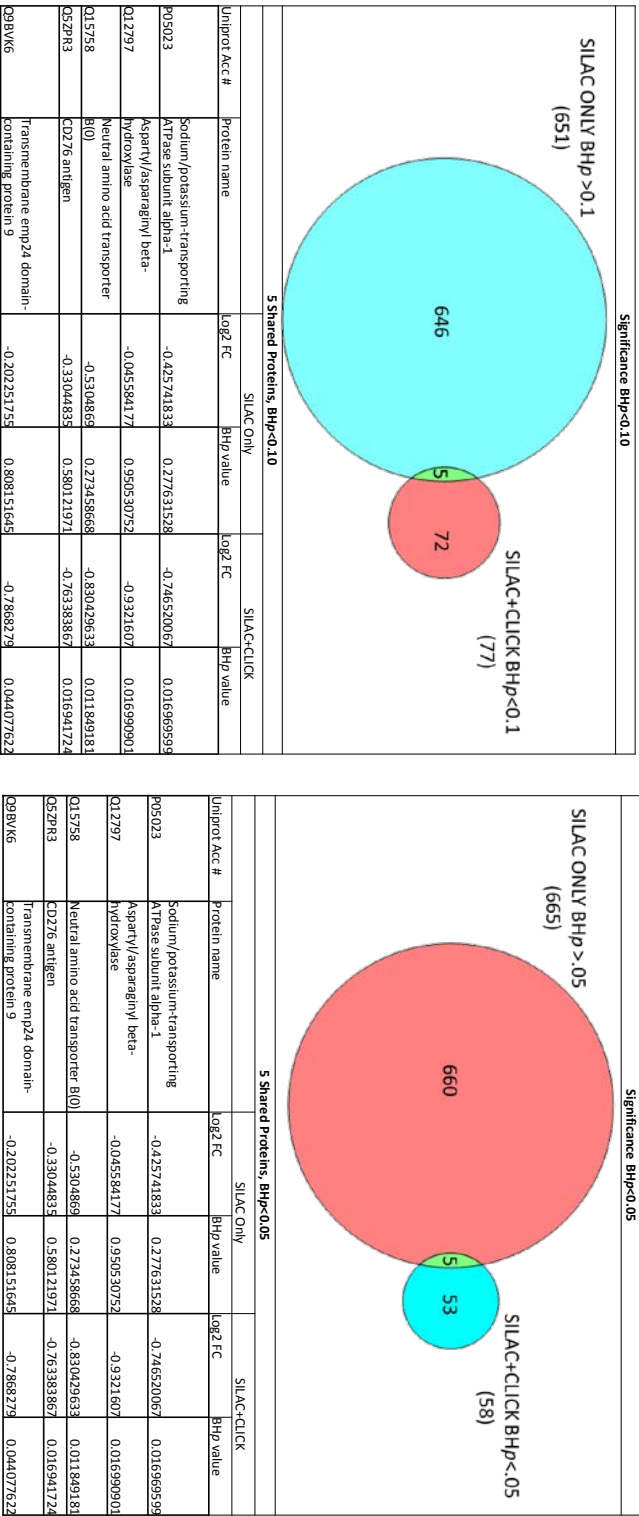
Several proteins potentially undergoing differential glycosylation following the induction of TGF $\beta$  were identified from the label-free quantification experiments. We next wanted to ascertain whether changes in underlying protein expression were contributing to the differential capture quantity. Therefore, SILAC metabolic labeling, for relative protein quantification, was included in the workflow prior to ManNAcAz labeling and gene induction. A total of 700 cell lysate-derived proteins were quantified in three biological replicates of SILAC and ManNAcAz labeled H358 TGF $\beta$  and EV induced cells without alkyne-agarose bead enrichment (“SILAC-Only”) (**Table 12,**

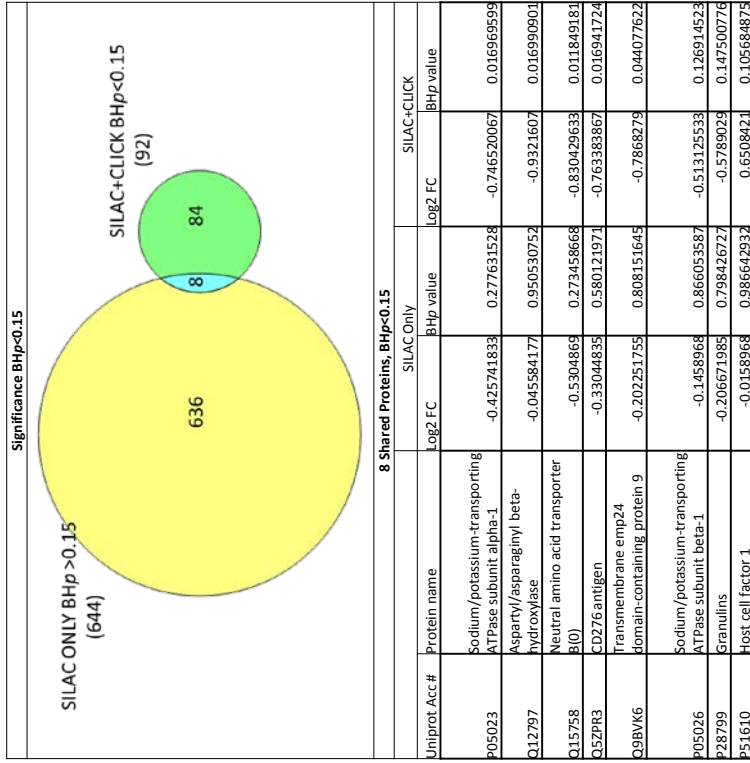
**suppl. materials**). Of the 700 quantified proteins, 11% were identified as glycoproteins. For the equivalent three replicates which underwent the alkyne-agarose bead enrichment (“SILAC+Click”), 392 total proteins were quantified; 76% of which matched the UniProt glycoprotein keyword search (**Table 13, suppl. materials**). Comparison of proteins identified with and without enrichment revealed a total of 91 shared proteins between experiments (**Table 14, suppl. materials**). We next wanted to identify which shared proteins were significantly changing in the SILAC+Click experiments while not significantly changing in the SILAC-Only experiments. Three Benjamini-Hockberg corrected  $p$ -value cutoffs ( $BH_p = 0.05, 0.10, \text{ and } 0.15$ ) were applied to the averaged  $\log_2$  fold-change of the SILAC ratios (**Table 15**). At  $BH_p = 0.05$ , 665 SILAC-Only and 58 SILAC+Click proteins met this criteria. There were 5 proteins shared in this assessment, including CD276, Transmembrane emp24 domain-containing protein 9, Neutral amino acid transporter B, Aspartyl/asparaginyl beta-hydroxylase, and Sodium/potassium-transporting ATPase subunit alpha-1. For  $BH_p = 0.10$ , there were 651 SILAC-Only and 77 SILAC+Click proteins; the same 5 proteins were identified in the overlap between groups. At  $BH_p = 0.15$  cutoff, 644 SILAC-Only and 92 SILAC+Click proteins remained. In addition to the 5 proteins shared from the lower  $BH_p$  values, Host cell factor-1, Granulins, and Sodium/potassium-transporting ATPase subunit beta-1 fit our criteria at this significance level.

An assessment of proteins significantly changing in both SILAC+Click and SILAC-Only experiments was also carried out. Two proteins, Integrin beta-4 and Transforming growth factor-beta-induced protein ig-h3, met this criteria at  $BH_p < 0.05$ . In



**Table 15: Shared Proteins Significantly Different in SILAC+Click and Not Significantly Different in SILAC-Only Fraction at 3 Benjamini-Hockberg P-value Cutoffs**



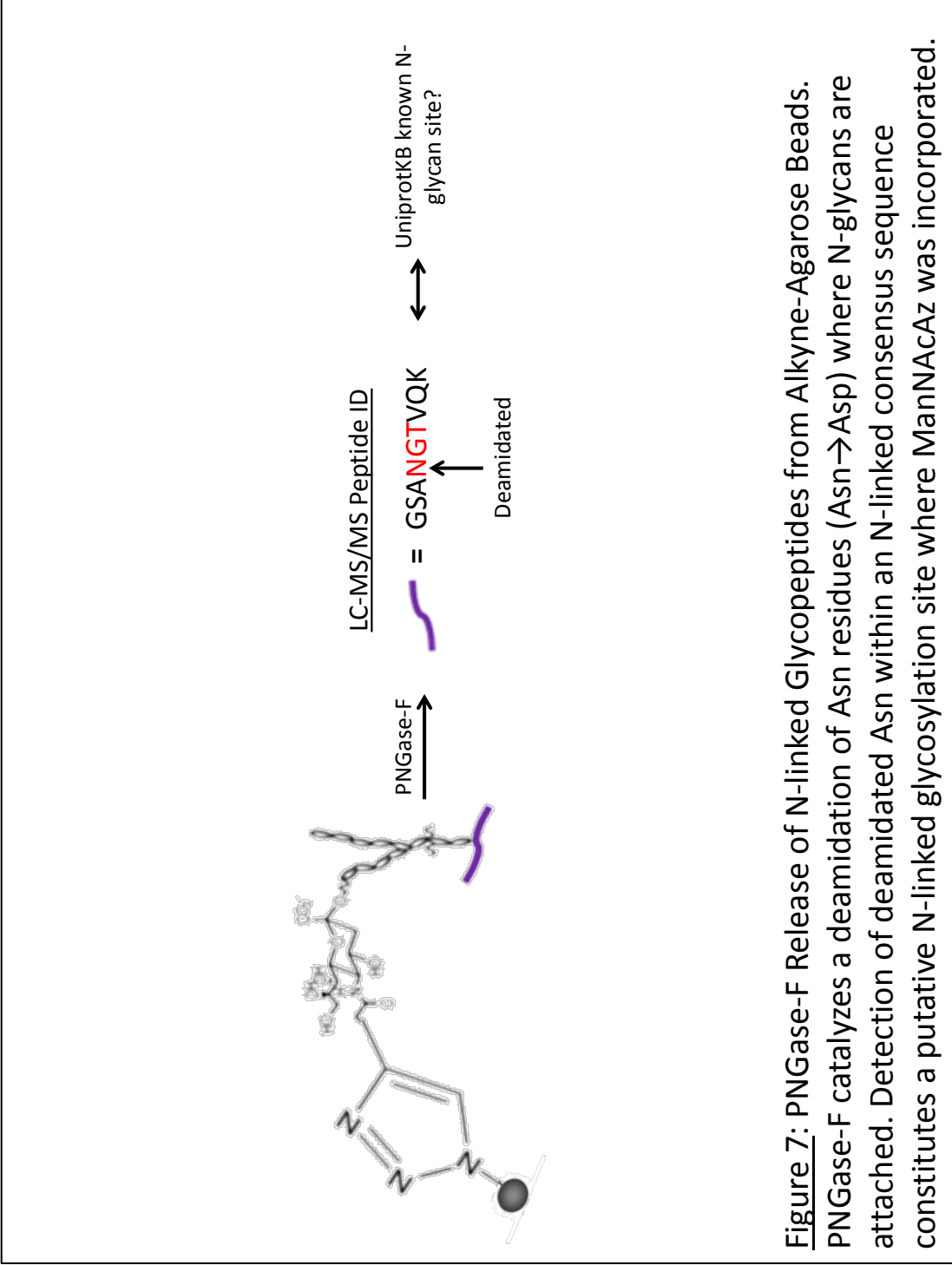


an equivalent examination at  $BHp < 0.10$  and  $BHp < 0.15$ , a third protein, Myoferlin, was also significant in both groups.

Finally, the 58 SILAC+Click proteins captured at significantly different levels ( $BHp < 0.05$ ) were submitted for pathway analysis (**Table 13**). Enriched pathways represented by these proteins included: Metabolic processes (5 proteins), cell adhesion molecules (4 proteins), ECM-receptor interaction (2 proteins), Mucin type O-glycan biosynthesis (2 proteins), Folate biosynthesis (2 proteins), malaria (2 proteins), p53 signaling pathway (2 proteins), complement and coagulation cascades (2 proteins), and phagosome (2 proteins).

### 3.3.3 PNGase-F Release of N-linked Glycopeptides

The previously-trypsin digested alkyne-agarose beads from our SILAC+Click experiments were retained to assess N-linked glycosylation sites where ManNAcAz was being incorporated (**Figure 7**). Following the removal of any carryover tryptic peptides, the beads were incubated with PNGase-F and evaluated the peptides by LC-MS/MS. We observed a total of 553 deamidated (Asn or Gln) peptides representing 224 unique proteins, of which 77% matched the UniProt glycoprotein keyword search. Peptide data was filtered to include only those with a deamidated Asn (411 peptides) and at least one N-linked glycan consensus sequence (N-X-S/T), narrowing the list to 385 peptides (**Table 16, suppl. materials**). From these 385 peptides, 170 unique proteins were represented, with 91% matching the glycoprotein keyword. We next cross-referenced a randomly selected subset of 55 observed peptides against the known glycosylation sites in the UniProt knowledge Base. In this selected set, 7 of the peptides were discovered to be overlapping in identifying specific N-linked sites, differing only in where N- or C-



**Figure 7:** PNGase-F Release of N-linked Glycopeptides from Alkyne-Agarose Beads. PNGase-F catalyzes a deamidation of Asn residues (Asn→Asp) where N-glycans are attached. Detection of deamidated Asn within an N-linked consensus sequence constitutes a putative N-linked glycosylation site where ManNAcAz was incorporated.

terminal trypsin cleavage had occurred. Furthermore, 3 peptides were deamidated at Asn residues outside of their N-linked consensus sequence and 3 were observed having multiple deamidated Asn in their sequence (both within and outside of consensus sequences). A total of 43 of the 55 peptides had deamidation localizations which matched documented UniProt KB sites of glycosylation and 3 sites were identified in non-annotated N-linked consensus sequences (**Table 17, suppl. materials**).

In addition to sequence assessments, a comparison of the representative proteins from PNGase-F released peptides to those of the equivalent samples digested with trypsin (SILAC+Click experiments) was performed (**Table 18, suppl. materials**). A 63% overlap of proteins derived from PNGase-F peptide release was observed in the trypsin data. Glycoprotein content was 51%, 67%, and 92% in the identifications unique to PNGase-F, unique to trypsin, and the shared proteins, respectively.

### 3.3.4 H358 Global N-linked Glycomics

Global changes in the N-linked glycome of H358 cells induced or non-induced to express of TGF $\beta$  or EV were screened. Cell lysate-derived N-glycans were released by PNGase-F and prepared by one of two methods of derivatization prior to MALDI-FT-ICR analysis: Permethylation or Linkage-specific sialic acid esterification. For the Permethylated glycans, evaluation was initially carried out in the 500-5000 m/z mass range from PNGase-F digested or no-enzyme controls (**Figure 8**). Obvious differences in the distribution of spectra from PNGase-F digested samples relative to no PNGase-F controls were observed; notably the presence of peaks above 2000 m/z in treated samples. We next compared the spectra of induced TGF $\beta$  and EV in the 1150-5000 mass range (**Figure 9**). Several shared structures were identified, including oligomannose and bi-,

tri-, and tetra- antennary complex structures between the TGF $\beta$  and EV induced samples. Notably, one species, Hex6HexNAc5 (m/z = 2519.3), was observed only in the TGF $\beta$ -induced sample. In addition, mono- and tri-sialylated complex glycans were noted in the upper mass ranges of both samples. For the linkage-specific sialic acid glycan analysis, spectra were assessed from 1150-2600m/z (**Figure 10**). Several of the low-mass oligomannose species, including Hex6HexNAc2 through Hex9HexNAc2 shared identification in the permethylated data, including both TGF $\beta$  and EV groups. The anomeric configuration of sialic acid on seven different bi-antennary complex glycans was resolved. These included Neu5Ac $\alpha$ 2-6 Hex5dHex1HexNAc4 (m/z = 2128.7) and Neu5Ac $\alpha$ 2-6Hex5dHex2HexNAc4 (m/z = 2275.8), which were also noted in the permethylated spectra. As a general observation,  $\alpha$ 2-6 linkages were predominant on mono-sialylated species, while di-sialylated glycans displayed a combination of  $\alpha$ 2-3 and  $\alpha$ 2-6. Relative peak abundance suggests the major species present in TGF $\beta$  and EV induced H358 cell lysate were Neu5Ac $\alpha$ 2-3Neu5Ac $\alpha$ 2-6Hex5dHex1HexNAc4 (m/z 2402.8). However, a clear increase in the proportion of Neu5Ac $\alpha$ 2-6Hex5dHex1HexNAc4 (m/z 2128.7) relative to m/z 2402.8 is evident in the TGF $\beta$ -induced sample.

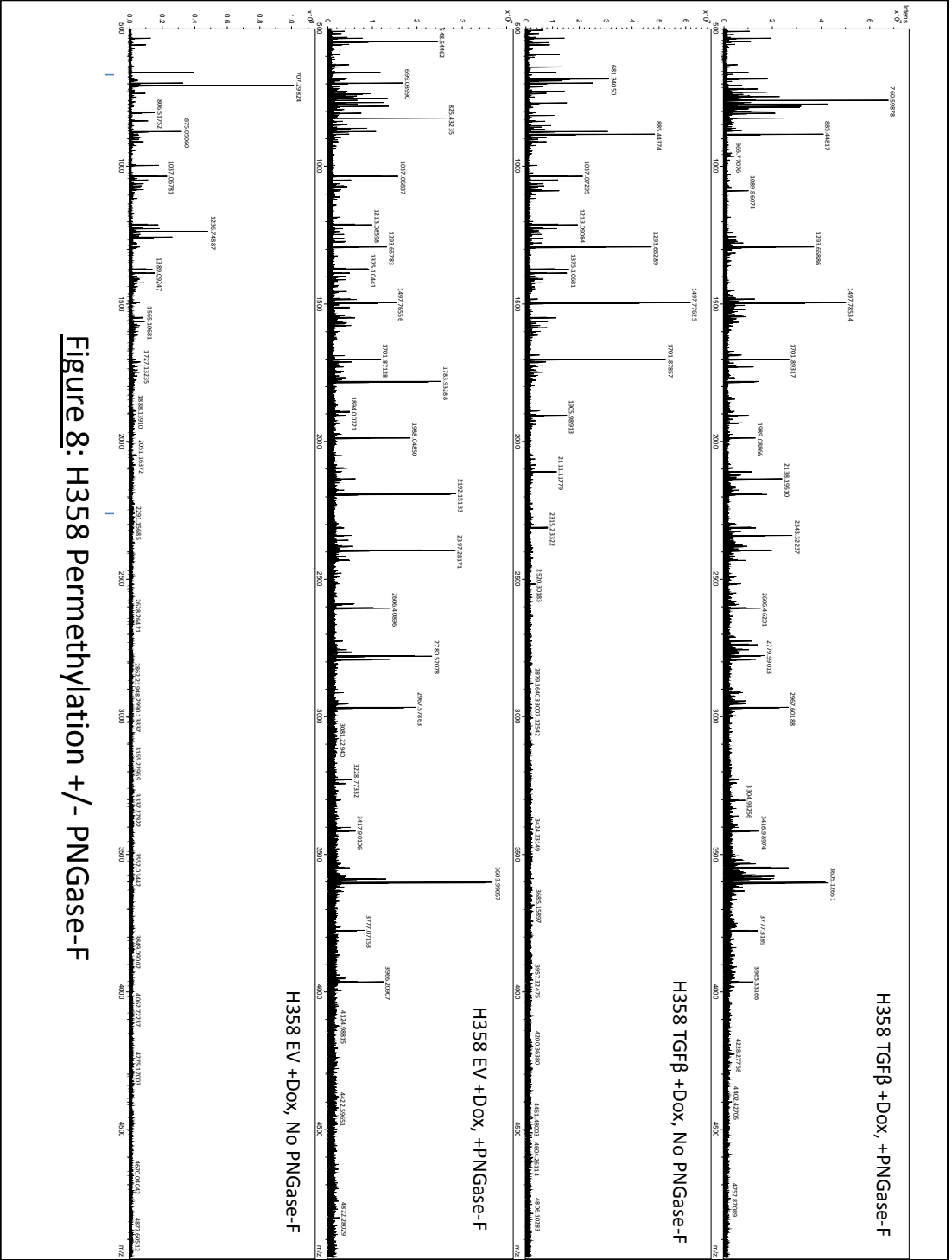
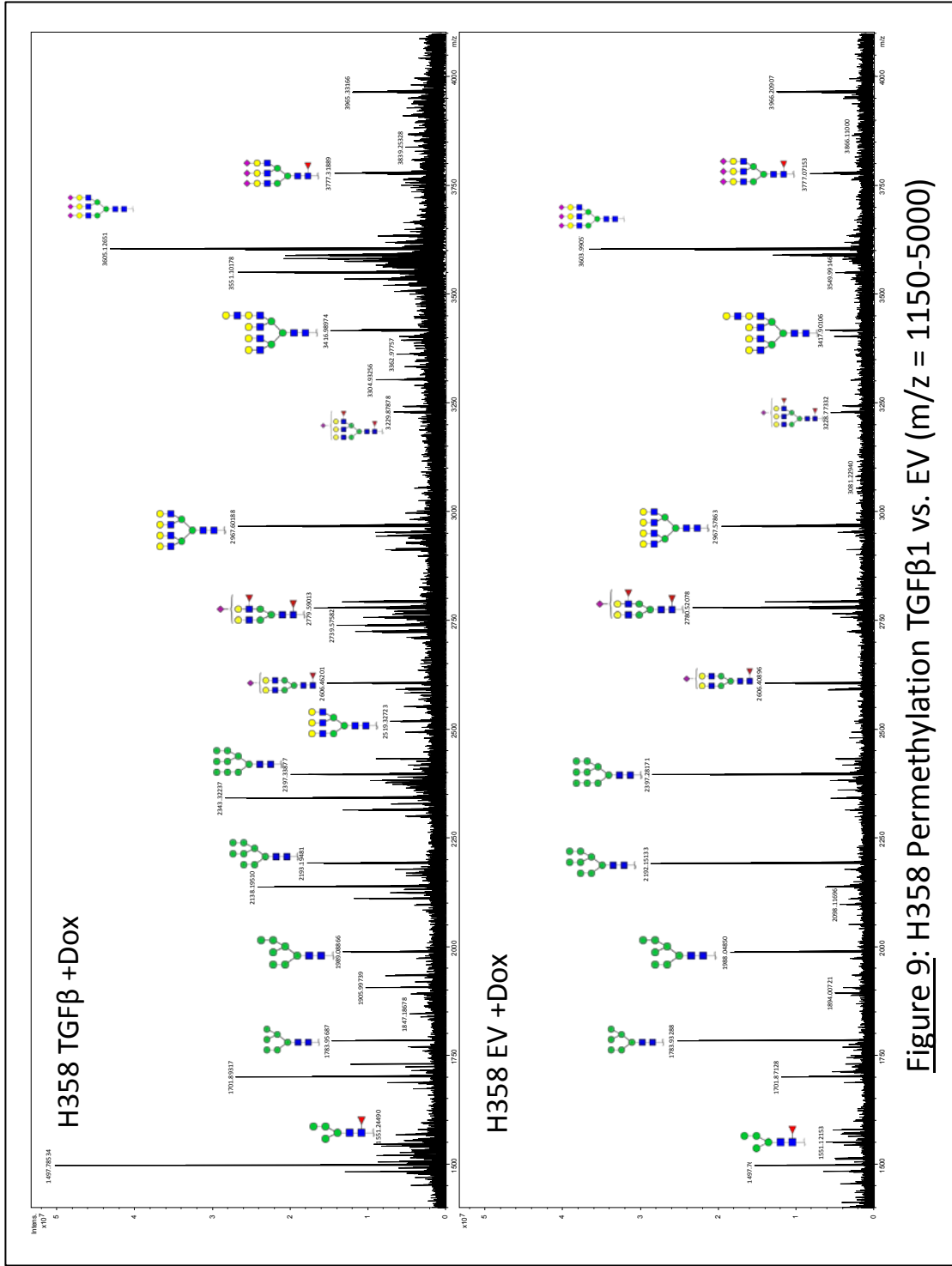
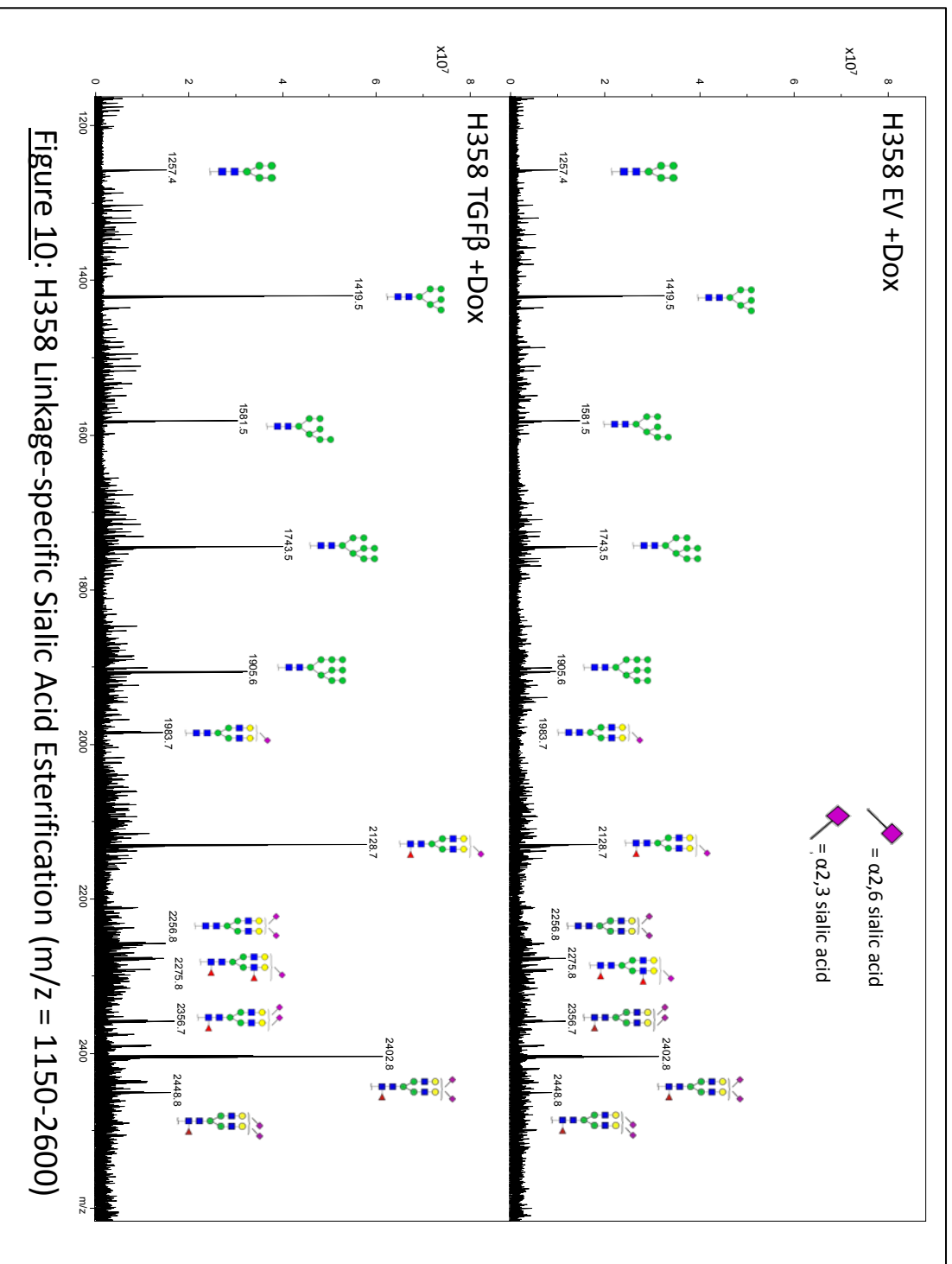


Figure 8: H358 Permethylation +/- PNGase-F







**Figure 10: H358 Linkage-specific Sialic Acid Esterification (m/z = 1150-2600)**

### 3.4 Discussion

Our initial experiments evaluated glycoprotein quantities following azido-sugar metabolic labeling and alkyne-agarose bead capture of H358 cell lysate following the induction of TGF $\beta$ , Zeb, or EV. We compared experimental groups by LC-MS/MS and label-free quantification by calculating the average log<sub>2</sub> fold change of nSCs and applying a moderated t-test to deem significance ( $p < 0.05$ ). Of note, the induction of TGF $\beta$  in ManNAcAz-labeled cells resulted in significant differences of the captured quantity of several glycoproteins relevant to processes such as cell adhesion, cytoskeletal regulation, ECM-receptor interaction, N- and O-glycan biosynthesis, solute transport, and metabolic pathways (**Tables 5 -7**). For example, relative to Zeb, TGF $\beta$ -induction resulted in the increased capture of Macrosialin, Cub-domain containing protein-1, Integrin alpha-2, and Integrin beta-6. In addition, differences were apparent in several glycosyl transferase enzymes, including GALNT10, MGAT2, B4GALT1, and GALNT6, in the same comparison. Furthermore, the TGF $\beta$ 1 ligand, which harbors three N-linked glycosylation sites, was the most significantly increased glycoprotein captured in all comparisons where its expression was induced. Experiments utilizing the GalNAcAz label with TGF $\beta$  induction identified significant differences in known O-linked glycoproteins such as SLC31A1 copper transporter and CD44. Moreover, the cell membrane-associated glycoprotein Basal cell adhesion molecule and cytoskeletal organization protein PDZ and LIM domain protein-7 were captured at altered levels with GalNAcAz labeling. In the final two comparisons, based on ManNAcAz labeling and Zeb-induction, we observed decreased capture of cell adhesion proteins (Intercellular adhesion molecular-1, Cadherin-3, Tetraspanin-13, Dystroglycan, Syndecan-1), ECM-

based proteins (Thrombospondin-1 and Laminin), and cell surface receptors (Plexin-B2, Receptor tyrosine protein phosphatase alpha and Epidermal growth factor receptor).

Further examination of data from comparisons where the Zeb gene was induced (**Tables 7, 10 and 11**) revealed a trend suggesting a bias towards lower nSC values for this experimental group. A review of the data for all of the individual samples indicates that this may be due to changes in the quality of MS acquisitions over the 14 month time period in which these experiments were carried out. This assertion is based on the total acquired spectra and normalization factor for each sample provided by the ProteoIQ software used for nSC generation (**Table 3**; rows 3 and 4). For example, within experiments performed at individual time-points (as indicated by the dates and colors on **table 3** row 5), the total spectra and normalization factors are relatively consistent. However, large variations in these values are apparent when equivalent samples acquired at different time-points are compared. In this regard, it is quite possible that the quality of online chromatographic separations varied over time because analyses were performed on a highly-operated core-facility mass spectrometer. We were prepared for this possibility and addressed it in the design of our data analysis. For each comparison (**Tables 5-11**), the equivalent replicates for each group were selected from experiments carried out at the same time-point (see **Table 3**: row 1). For example, if group A was being compared to group B, replicate 1 for group A was selected from the same experiment as replicate 1 for group B; the same being true for replicate 2 and replicate 3. This approach accounts for variability in instrument performance and alleviates any bias which would have been introduced from using data acquired over a long period. For our goal of initially assessing azido-sugar labeling, enrichment, and LC-MS/MS

quantification, this strategy was an appropriate choice for a rapid, preliminary analysis prior to implementing a more reliable quantification method.

The experimental groups studied in our label-free quantification served to establish the feasibility of our experimental procedures prior to double metabolic labeling. Two concentrations of doxycycline for induction (data not shown), two azido sugars (ManNAcAz and GalNAcAz), and samples derived from both cell lysate and secreted protein fractions were evaluated. In addition, glycoprotein capture following the induction of the gene *Zeb*, as well as non-induced controls, was assessed. We were able to identify significant differences in captured glycoproteins using label-free quantification and utilized this data as a guide for selection of TGF $\beta$ 1 and EV experimental groups only in our double metabolic labeling experiments.

Next, protein level changes were gauged to determine if this was contributing to differential capture of azido-sugar labeled glycoproteins. SILAC relative protein quantification was used in these experiments, rather than label-free quantification, because of the increased reliability of data associated with this method(215). Three biological replicates of SILAC and ManNAcAz metabolic labeling were performed on TGF $\beta$  and EV- induced H358 cells and proteins were quantified both before (SILAC-Only) and after alkyne-agarose bead capture (SILAC-Click). In order for this strategy to have merit, candidate proteins must be quantified in both of these fractions. Therefore a two-dimensional chromatographic separation (SCX and Reverse phase) of peptides was applied to maximize the number of identifications made in both groups. As a result, 700 and 392 proteins were quantified in the SILAC-Only and SILAC+Click groups (respectively), however a relatively small subset of proteins were shared between the

fractions (91 total). To investigate this low overlap, a GO Analysis was performed on total proteins quantified in each fraction (data not shown). The SILAC-Only group contained a diverse set of proteins representative of broad biological processes derived from intercellular, macromolecular complexes, and organelle-based components. Conversely, the majority of quantified hits from the SILAC+Click fraction stemmed from membrane-based and cell surface components. In addition, there was an obvious discrepancy in the proportion of glycoproteins identified for each fraction (11% v s. 76%). Therefore, the lack of overlap between SILAC-Only and SILAC-Click may be explained by the efficiency with which the Click chemistry based enrichment selectively pulls down azido-sugar incorporated-glycoproteins. A more prudent approach to assess protein quantity prior to enrichment would be Western blot analysis. Candidates captured by alkyne-agarose beads at significantly different levels following TGF $\beta$ 1 induction, with relevant biological functions to invasive phenotypes, could be specifically assayed for this purpose without the necessity of time consuming LC-MS/MS acquisitions and bioinformatics analyses.

For statistical analysis of our double metabolic labeling, a moderated t-test was applied to the averaged log<sub>2</sub> fold change of the SILAC ratios with respect to TGF $\beta$ 1 induction. Assessment of significance at BH $p$ -values of 0.05, 0.10, and 0.15 was carried out to evaluate shared proteins which were significantly changing in SILAC+Click while not reaching significance in SILAC-Only. The same five proteins were observed meeting this criteria at both BH $p$  = 0.05 and BH $p$  = 0.10; however, an additional 3 proteins were included at BH $p$  = 0.15. A UniProt search of these proteins indicated that seven of the eight are known glycoproteins; with only Sodium/potassium-transporting ATPase subunit

alpha-1 not meeting the designation. Interestingly, the beta-1 subunit of this protein also fit inclusion criteria at  $BHp = 0.15$ , where it was observed to be decreasing about 30% following  $TGF\beta 1$  induction and alkyne-agarose bead capture (**Table 15**). The Na/K ATPase is composed of heterodimeric alpha and beta subunits; where alpha carries out the catalytic role and beta functions in regulating plasma membrane localization of the pump. Recently, a role for the beta-1 subunit in epithelial cell polarity and cell-cell adhesion was identified and shown to be modulated by N-linked glycosylation (216,217). The authors conclude that hypo-glycosylated beta-1 subunits resulted in decrease cell-cell contacts through modification of the ATPase's stability within the adherens junction. Regarding the detection of the alpha-1 subunit in our SILAC+Click fractions, it is plausible that the intact, heterodimeric form of the ATPase was captured through ManNAcAz incorporation within the beta-1 subunit. Consistent with this is the absence of sequenced peptides identified from the alpha subunit in our PNGase-F mediated release of glycopeptides from the alkyne-agarose beads; whereas two unique beta-1 peptides, covering two N-linked glycosylation sites, were identified (**Table 16**) .

Another putative candidate identified from this work was CD276, which was captured at significantly decreased levels following the induction of  $TGF\beta$ . This membrane-based-glycoprotein, also known as the T-cell B7-H3 co-stimulatory molecule, has previously been implicated in the progression of cancer through effects on anti-tumor immunity(218,219). However, very little information exists regarding the biological role of its 6 potential N-linked glycosylation sites. One recent methods-based study uncovered two sites of altered sialylation in CD276 stemming from hepatocellular carcinoma cells treated with doxorubicin(220). Another study indicates that the B7 family of co-

stimulatory molecules are constitutively expressed on airway epithelial cells and may contribute to increased immune surveillance and inflammation at these surfaces(221). Unfortunately, an assessment targeting the sialylation status of CD276 in lung cancer has yet to be carried out.

The aspartyl/asparaginyl beta-hydroxylase (ASPH) was observed having the lowest relative change in protein expression following TGF $\beta$  induction while simultaneously having the largest significant difference following ManNAcAz capture (47% decrease following TGF $\beta$ 1 induction). This glycoprotein is found in the ER plasma membrane where it functions in the hydroxylation of EGF-like domains of transient proteins. The expression of ASPH has previously been noted as up-regulated at the mRNA level in several tumor cell lines and has therefore been suggested for further scrutiny as a putative biomarker(222). A UniProt search of ASPH indicates that it has two predicted N-linked glycosylation sites based on its amino acid sequence, however no published reports have confirmed this. In addition, Granulins was also identified from double metabolic labeling experiments as a possible candidate of altered sialylation. This secreted protein has functions in wound healing/tissue repair, inflammation, embryogenesis and has been implicated in the progression of multiple cancers (223,224). There are 5 predicted sites of N-linked glycosylation on granulins and a recent publication has characterized 4 of them(225). Interestingly, although the authors noted the presence of several different fucosylated complex N-glycans, none of the species were sialyated.

The lack of protein-level data on the remainder of the significantly different SILAC+Click proteins does not preclude further investigation of these candidates. KEGG

analysis of these 58 proteins identified that they were enriched in pathways related to metabolic processes, cell adhesion, extracellular matrix-receptor interaction, and other pathways with relevance to the acquisition of invasive tumor phenotypes. Targeted studies of each individual species may therefore be warranted. For example, immunoprecipitation could be used to attain more-purified starting material for intact glycopeptide analysis or glycomic studies. This approach would be contingent on the availability of highly-specific antibodies for each target and MS instrumentation capable of these types of acquisitions. These methods, however, have the potential for providing detailed characterization of glycan structures and glycosylation site occupancy information which could then be assessed as differentiating characteristics, alone or in combination with other parameters, for biomarker development.

Following the double metabolic labeling experiments, we next investigated the value of LC-MS/MS analysis of PNGase-F released glycopeptides from previously trypsin-digested alkyne-agarose beads. This approach was used to specifically survey the N-linked glycopeptides which incorporated ManNAcAz; as a means to potentially identify the sites where this was occurring. The principle of these experiments was based on the enzymatic specificity of PNGase-F, which cleaves the linkage between the core GlcNAc and the asparagine residue of all N-glycans (with the exception of those bearing a core ( $\alpha$ 1,3) linked fucose). During this reaction, the asparagine to which the glycan was attached is deamidated (converting it to Aspartic Acid), resulting in a 1Da mass shift detectable by MS. This approach to glycosylation site identification has previously been applied to N-glycan enrichment strategies including hydrazide capture and lectin affinity



chromatography, but never utilizing Click-chemistry-based alkyne bead enrichment (179,180,226).

Over 500 deamidated peptides were initially identified from the PNGase experiments. Sequential data filtering, from all deamidated peptides (Asn + Glu) → deamidated Asn only → deamidated Asn only and at least one N-linked consensus sequence, was applied to reduce the list to the most probable group for harboring N-glycans. This process brought the number of peptides down to 385, representing 170 unique proteins (91% glycoproteins). A subset of these peptides were selected for cross-reference to assess how our putative site localizations compared to those on UniProtKB. This approach was selected because of the large list of peptides and the requirement of manual annotation. We discovered that, for some individual proteins represented by multiple peptides in our list, the same deamidation site was being identified. This phenomenon appears to be a consequence of slight variation in where N- or C-terminal trypsin cleavage occurred. As expected, deamidated Asn outside of N-X-S/T sequences were identified for a few of the peptides, reiterating the potential for false site identifications due to random deamidations(227). Modern approaches to circumvent this are focused on performing PNGase-F digest in the presence of isotopically-labeled water for the incorporation of <sup>18</sup>O into Asp following glycan release. This method, termed Isotope-coded glycosylation site specific tagging (IGOT) has demonstrated utility for increasing the fidelity of putative N-linked sites(228). We also evaluated how deamidation sites from our selected list were annotated in UniProtKB. This database makes use of two designations for the assignment of glycosylation: Published or Sequence analysis. Published sites are derived from experimental observations while

sequence analysis are predicted sites based on the presence of an extracellular consensus sequence and confirmation from an external prediction service such as NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>). Of the 43 non-repetitious peptides queried in our selected list, 24 matched UniProt listed sites by sequence analysis, highlighting the lack of existing information on N-linked glycan localizations. Therefore experimental approaches capable of cataloguing these sites will be useful for creating libraries from which experimental data may be compared.

An additional comparison of the protein identifications made following PNGase-F release to those discovered from the original trypsin digest was performed. Approximately 63% of the total proteins represented in the PNGase-F based experiments had already been observed in the Trypsin fraction. Several potentially relevant glycoproteins were uniquely identified with either enzyme; however this may be the result of non-equivalent chromatographic separation and/or data processing steps used for either analysis. For example, the trypsin based experiments were carried out using a 5-step 2D-LC chromatographic separation prior to LC-MS/MS, while only a single dimension was used for PNGase-F release acquisitions. The 392 trypsin-based identifications were made based on the application of our quantification criteria; which included the requirement of  $\geq 2$  SILAC ratios from the 3 biological replicates. In addition, trypsin data was not limited to deamidated peptides only, whereas PNGase-F data was immediately filtered to include only peptides with this variable modification. Therefore a more even gauge of protein content might be attained by the application of standardized chromatographic separation and data processing parameters prior to comparison.

To provide a more comprehensive analysis of the glycosylation profile of the H358 cells induced to express TGF $\beta$  or EV, we explored the N-linked glycoforms detectable by PNGase-F release and MALDI-FT-ICR analysis. Two experiments were performed, each utilizing a different glycan derivatization strategy. Permethylation was carried out to optimize FT-ICR settings for the detection of sialic acid-bearing moieties and linkage-specific sialic acid esterification to increase the detail of identified structures. From this cumulative effort, a wide range of N-glycans were acquired, including oligomannose, bi- tri- and tetra- antennary complex glycans, mono- and tri-sialylated, and core-fucosylated species. We observed one glycan which was detected only in the permethylated TGF $\beta$ -induced group (**figure 9**,  $m/z = 2519.3$ ). Furthermore, inferences regarding relative glycan abundance led us to discover a disproportionate increase one of the sialylated N-glycans in TGF $\beta$  relative to EV in the linkage-specific derivatization (**figure 10**,  $m/z = 2128.7$ ).

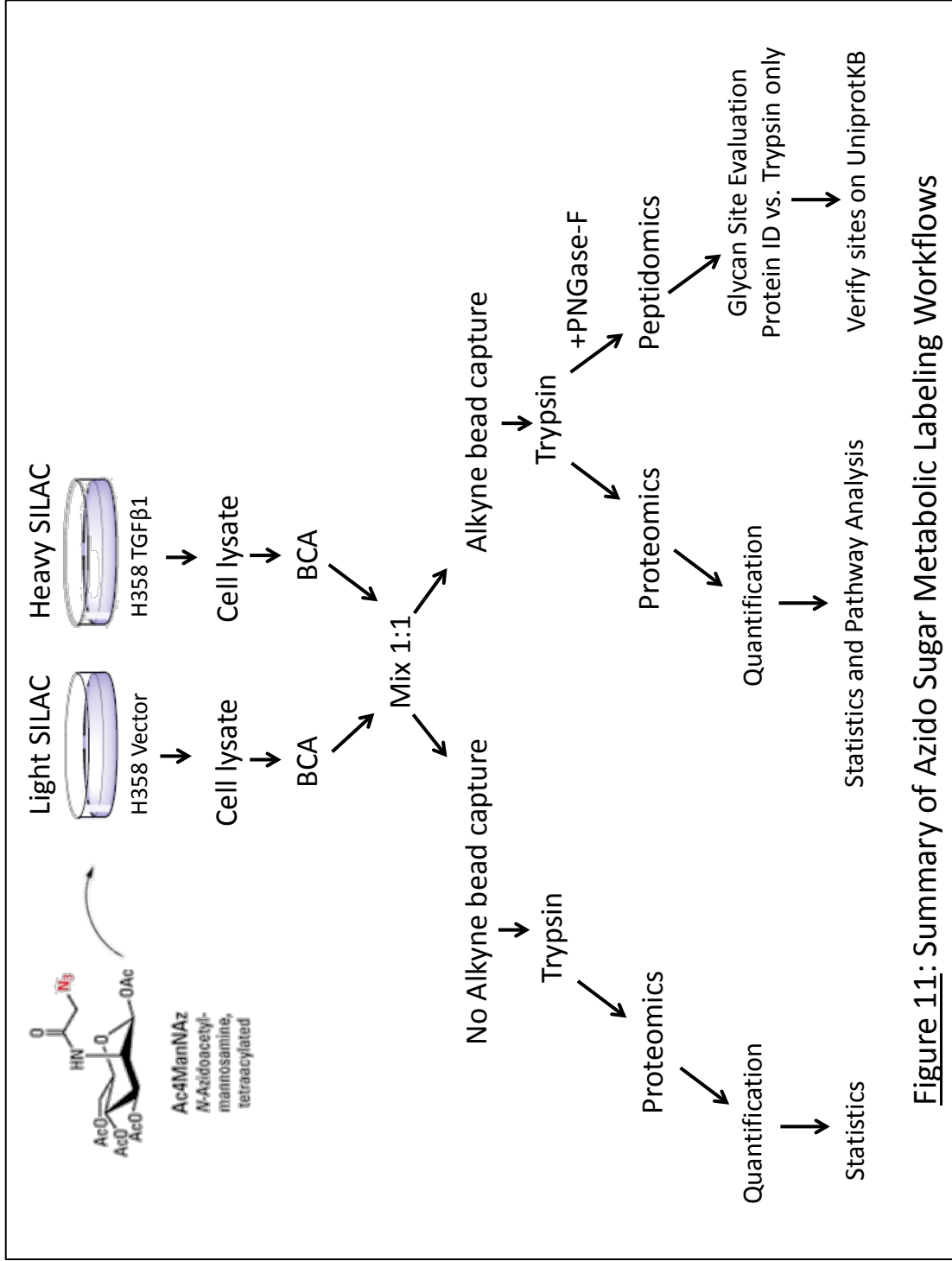
One of the limiting factors to our H358 total N-glycome data is that only one replicate for each derivatization method was examined. This was primarily due to excessive difficulties in attaining high-intensity spectra with the H358 cell lysate. For the linkage specific work, two protein precipitation methods were tested, two protein concentrations for PNGase-F digest, and serial cotton tip HILIC enrichment prior to attaining the presented spectra. Despite lack of demonstrated reproducibility, we believe that the application optimized methods used here will be beneficial for future glycoproteomic biomarker discovery endeavors. For example, by utilizing two distinct derivatization methods we were able to identify a more diverse set of glycans than would have been evident from either preparation alone. Permethylation data identified two tetra-

antennary and three tri-antennary species which were not observed in the linkage-specific experiment. At the same time, four distinct bi-antennary complex glycans, and their sialic acid linkages, were discerned exclusively in the esterification fraction.

Finally, the use of high-resolution/high mass accuracy MS analysis greatly facilitated the structural characterization of these glycans. This acquisition strategy is based on the sensitivity of the FT-ICR mass analyzer for resolving the individual monosaccharide constituents of a given glycan moiety(230). Importantly, this type of mass analyzer has demonstrated utility for use with various ionization techniques including ESI, MALDI, and fast-atom bombardment. Thus, FT-ICR analysis of N-linked glycans is an ideal platform for rapid, sensitive studies which can be carried out on vast array of MS instrumentation.

In summary, we demonstrated approaches for evaluating glycoproteins incorporating an analog of sialic acid, glycoform structures, and the identification of putative N-linked glycosylation sites from the H358 NSCLC cells. A workflow was explored utilizing the incorporation of azido-sugars and gene induction followed by enrichment of glycoproteins bearing these analogs (**Figure 11**). LC-MS/MS and label-free quantification was applied to gain an initial look at targets which may be undergoing alternative glycosylation as a result of TGF $\beta$ 1 or Zeb induction. The approach was narrowed to assess only TGF $\beta$ -influenced alterations in sialylated proteins and included SILAC metabolic labeling. In these experiments, relative protein quantity was calculated both before and after alkyne agarose bead enrichment so that inferences could be made regarding how protein expression levels were affecting captured targets. Further utility from the alkyne-bead capture format of these experiments was gained by performing a

secondary digest, with PNGase-F, to identify site localizations of where ManNAcAz was being incorporated. Finally, a preliminary analysis of the global N-glycome of the H358 was carried out. Two separate derivatization methods were implemented and glycans were resolved by high-resolution/high mass accuracy MALDI-FT-ICR to ascertain if TGF $\beta$ 1 was influencing the glycoforms and/or types of sialic linkages present. The combined use of these methods provided multi-faceted insight as to how TGF $\beta$ 1 influenced the glycoproteome of this model and we therefore assert that these techniques have merit for future biomarker discovery endeavors.



**Figure 11:** Summary of Azido Sugar Metabolic Labeling Workflows

## Chapter 4

# Glycomic Assessment of Human NSCLC Biofluids and Tissue

## 4.1 Introduction

The initial stages of glycoproteomic biomarker discovery are most often carried out using materials derived from in vitro or animal models of disease, tissue lysates, or serum/plasma (14,231). Each discovery substrate has distinct advantages and disadvantages, therefore sample availability and experimental strategy are often deciding factors for selection. For example, plasma/serum matrices are easily obtainable and have established utility in clinical testing. However, these samples may not be ideal for the study of low-abundance proteins or post-translational modifications because of the need for extensive pre-analytical preparations to remove high-abundance proteins. Cell culture models provide a cheap, abundant source of materials and are compatible with metabolic-labeling to confidently quantify large numbers of analytes. However, these systems may not truly be representative of the in vivo tumor microenvironment, often restricting the applicability of findings. To circumvent these limitations, several approaches for biomarker discovery have shifted towards analyzing biofluids in direct contact with the tissue of interest (“Proximal fluids”)(15,232–234). These fluids contain tissue-specific leaked and secreted proteins at higher concentrations than found in serum. In addition, many proximal fluids are readily accessible (saliva, urine, tears) or routinely collected during diagnostic procedures (Bronchoalveolar lavage fluid (BAL), cerebrospinal fluid, gastric secretions). Furthermore, these types of samples are ideal for glycoproteomic or glycomic studies because of the frequency with which this PTM occurs along the secretory pathway.

Another useful strategy for biomarker discovery has been the analysis of tumor tissue-lysate(235). These studies bypass several of the limitations of other starting



materials because of their direct nature; without dilution of analytes in heterogeneous substrates. In the past, workflows for the preparation of tissue-derived samples required laborious processing steps and were incompatible with concurrent examinations to define histological features. More recently, methods have been developed for intact tissue profiling and spatial resolution of proteins, lipids, and glycan species(211,236,237). Importantly, these imaging mass spectrometry procedures can be used to define the distribution of various analytes in fresh-frozen or formalin-fixed paraffin embedded tissues and may be carried out in parallel to pathological examinations.

At the same time as biomarker efforts using proximal fluids and tissue-imaging approaches are moving forward, innovations in pre-analytical sample preparations and MS-instrumentation have come about. These advances have been integral in facilitating glycomic-based biomarker studies(230). For example, evolving derivatization methods have synergized with the development of high-resolution and mass accuracy MS systems to provide a robust means for the structural characterization of sialylated carbohydrate molecules. One such procedure, recently described by Reiding et al., encompasses chemical modification of sialic acid for the discrimination of anomeric linkages(212). When this method is combined with highly sensitive instrumentation, such as a MALDI-FT-ICR, a more detailed structural evaluation can be performed than previously possible. While the use of MALDI-FT-ICR is not novel for oligosaccharide analysis, implementation of this strategy for investigations based on proximal fluid derived glycans has not been previously investigated.

Therefore, we implemented a novel approach to the characterization of disease-specific glycan alterations in clinically-relevant biofluids using linkage-specific sialic

acid derivatization and MALDI-FT-ICR analysis. Pathologically-defined matched human samples, including BAL fluid and saliva, as well as fresh frozen lung tissue (tumor and adjacent normal) were obtained from the lung cancer bio-specimen repository network. Fluids were pooled and processed for derivatization and cotton-tip HILIC enrichment of N-linked glycans. Fresh frozen tissues were sectioned at 5µm thickness and processed on-slide prior to MALDI-FT-ICR acquisition of N-linked glycans; duplicate slides were prepared for histological evaluation by a board-certified pathologist.

## 4.2 Materials and Methods

### 4.2.1 Samples Description

10 de-identified, matched sets of human samples (5 benign, 3 squamous cell carcinoma, 2 adenocarcinoma) were obtained from the Lung Cancer Bio-specimen Resource Network (LCBRN). The included bio-specimens, and a link to the LCBRN website where the methods for collection are describe, can be found in **Table 19**. All samples were stored at -80°C until processing.

### 4.2.2 Bronchoalveolar Lavage Fluid Proteomics

#### 4.2.2.1 Sample preparation and In-solution digest

The 10 BAL samples were thawed on ice and 75µl aliquots were prepared as a working solution. 3µl of each samples was visualized by polyacrylamide gel electrophoresis (SDS-PAGE) (4-12% NuPage ®Bis-Tris Gel, Life Technologies) and silver stained (SilverQuest™, Life Technologies) to approximate total protein quantity. Normalized samples were brought up to 75µL total volume in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (Sigma), 10mM DTT, and allowed to reduce at 95°C for 10 minutes. After cooling to

room temperature, IAA was added to a final concentration of 20mM and alkylation proceeded for 30 minutes in the dark. Samples were digested at 37°C for 16 hours with 2µg proteomics grade trypsin. Following digestion, samples were desalted by µC18 Zip-Tip (Milipore), speedvac concentrated, and stored at -20°C until LC-MS/MS.

#### 4.2.2.2 LC-MS/MS and Bioinformatics

LC-MS/MS, Bioinformatics, and generation of normalized spectral counts for BAL samples were performed as described in section 3.2.5. Gene Ontology (GO) analysis of proteins uniquely identified in cancer-associated BAL fluid was performed using Web-based gene set analysis toolkit (Webgestalt: <http://bioinfo.vanderbilt.edu/webgestalt/>)(238,239). Protein identifications were compared to the normal distribution of genes in the human genome to identify biological processes and cellular components which were enriched ( $BH_p < 0.05$ ) relative to expected distribution. Specifications included: Organism = Homo sapiens, Id Type = uniprot/swissprot accession, Ref Set = entrezgene, Significance Level = Top10, Statistics Test = Hypergeometric, Multiple test correction = Benjamini Hockberg, Minimum number of genes = 2.

#### 4.2.3 N-linked Glycomics of Human Biofluids

Pooled BAL samples were prepared from each diagnostic group (Benign, SCC, AC) at 100µL total volume. Two saliva pools were prepared at 200µL total volume: 1. Benign- 40µL of each benign sample and, 2. Metastatic- 40µL from each of the metastatic groups (SCC+AC). Linkage-specific sialic acid esterification proceeded as described in section 3.2.10.1. MALDI-FT-ICR N-glycan analysis and bioinformatics were carried out as described in section 3.2.11.

## 4.2.4 N-linked Glycan Imaging of Lung Tissue

### 4.2.4.1 Sample preparation

Flash frozen lung tissues were prepared for MALDI-Imaging Mass Spectrometry (MALDI-IMS) by the formalin-fixed paraffin embedded method described by Powers et al.(240). Tissues were sectioned at 5 $\mu$ m and mounted onto positively charged glass slides compatible with the Bruker FT-ICR adaptor plate. Citraconic anhydride (Thermo) was used for antigen retrieval. Following buffer exchange and desiccation, .2 $\mu$ L of recombinant PNGase-F were applied uniformly using an ImagePrep spray station (Bruker Daltonics) and slides were incubated at 37°C for 2 hours in a humidified chamber. Slides were desiccated for 10 minutes and matrix (0.021g  $\alpha$ -Cyano-4-hydroxycinnamic acid (Sigma) in 3 ml 50% acetonitrile/50% water and 12 $\mu$ l 25% trifluoroacetic acid) was applied using the TM-sprayer™ (HTX Technologies). Duplicate slides of each tissue were prepared for hematoxylin and eosin staining, and definition of histological features was performed by a board-certified pathologist.

### 4.2.4.2 N-glycan MALDI Imaging Mass Spectrometry

Released N-glycans were acquired directly on tissue with a Solarix dual source 7T FT-ICR mass spectrometer ( $m/z$ = 500-5000). A SmartBeam II laser operating at 2000 Hz was used to acquire spectra across at a raster width of 50 $\mu$ m. Detected glycan species were viewed using FlexImaging 4.1 software. Identifications were made by searching spectra in the range of  $m/z$  = 1000-4000 against the glycan structures database of the Consortium for Functional Glycomics. Glycan structural annotations were produced in GlycoWorkbench 2.1.

Collection protocol for each specimen can be found at <http://avillage.web.virginia.edu/LungBio/standard-operating-procedures>

**Table 19: Lung Cancer Bio-specimen Resource Network Sample Types**

Diagnosis:	Benign				Squamous Cell Carcinoma (SCC)				Adenocarcinoma (AC)	
ID#:	W0081	W0099	V0034	W0005	W0086	V0229	V0232	V0233	V0231	W0216
Specimens:	Urine	Urine	Urine	Urine	Urine	Urine	Urine	Urine	Urine	Urine
	BAL	BAL	BAL	BAL	BAL	BAL	BAL	BAL	BAL	BAL
	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet	Cell pellet
	Serum	Serum	Serum	Serum	Serum	Serum	Serum	Serum	Serum	Serum
Paired Tissue:	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva
	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)	Lung (tumor)
	Lung (fibrosis)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)	Lung (normal)
	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)	Bronchus (normal)

BAL= Bronchoalveolar Lavage Fluid

## 4.3 Results

### 4.3.1 Bronchoalveolar Lavage Fluid Proteomics

We examined the proteomic contents of 10 human BAL fluids for an initial survey of constituents detectable by LC-MS/MS. These samples were derived from a pathologically defined set of matched biofluids and tissues, including 5x benign, 3x squamous cell carcinoma (SCC), and 2x adenocarcinoma (ADC). In addition, we gauged the relative abundance of identified proteins by label-free quantification to provide a well-annotated resource from which future studies may be compared.

Total protein abundance of each sample was estimated by SDS-PAGE and silver staining (**Figure 12**). Darker banding patterns were observed in the cancer relative to benign samples, with the exception of sample W0099. This visual assessment was used to gauge individual sample input for an in-solution trypsin digest and LC-MS/MS analysis of proteomic content. A total of 332 proteins were quantified from all samples following the application of post-processing parameters for the production of normalized spectral counts (**Table 20, suppl. materials**). Of the total proteins discovered, 35% matched the Uniprot glycoprotein keyword search. Grouping of the individual samples into two categories (Cancer and Benign) for comparison resulted in the identification of 25 proteins distinctly quantified in cancer-associated BALs and 12 unique to Benign (**Table 21, suppl. materials**). Glycoproteins identified only in the cancer group included Osteopontin, Clusterin, and Mucin-4. Conversely, Sialic acid synthase was quantified only in the benign group. For



**Figure 12:** Bronchoalveolar Lavage Fluid Protein Content. 3µl BAL fluid separated by 4-12% Bis-Tris PAGE and Silver-Stained.

shared proteins, discrepancies were obvious in the nSC quantity of cell adhesion and extracellular glycoproteins including an increase in CD59, Galectin-3 binding protein, Mucin-5, and Mesothelin, as well as decreases in the redox-related proteins Catalase, Peroxiredoxin-6, and Thioredoxin reductase-1, in the cancer group relative to benign. GO analysis of proteins uniquely identified from cancer BALs revealed significant enrichment of intracellular non-membrane-bounded, extracellular, and cytoskeletal cellular components, amongst others (**Table 22**). In addition, biological processes represented included protein-DNA complex assembly, chromatin assembly, and nucleosome organization.

#### 4.3.2 N-linked Glycomics of Human Bio-fluids

Matched human biofluids, including BAL and saliva, were processed for N-linked glycan profiling by linkage-specific sialic acid esterification and FT-ICR analysis. Total acquired spectra from each BAL pool (Bn, SCC, and ADC) were first evaluated from 1200- 3250 m/z (**Figure 13**). We observed abundant spectra ( $>10^8$  intensity) in each pool and therefore annotated individual peaks to assess differences between groups (**Figures 14-16**). A peak at m/z = 2301.8, correlating to a di- $\alpha$ 2-6-sialylated bi-antennary complex glycan, was the most prevalent species observed in all BAL samples. Within the benign pool, there was an increased abundance of Hex9HexNAc2 (m/z = 1905.7), Hex5dHex2HexNAc4 (m/z = 1955.7), and Hex5dHex1HexNAc4 (m/z = 1809.7) relative to SCC and ADC (**Figure 14**). The ADC pool had a distinct increase in an  $\alpha$ 2-6-mono-sialylated bi-antennary complex glycan (**Figure 14**: m/z = 1982.8) as well as an apparent decrease in the relative peak height of two di-sialylated bi-antennary complex glycans (**Figure 15**: m/z = 2355.9 and 2401.9).



**Table 22: Gene Ontology Analysis of Proteins Unique to Cancer Bronchoalveolar Lavage Fluid Samples**

Cellular Component					
Category	ID	# Genes in category	# Expected	# of genes matched	adjP value
extracellular space	GO:0005615	856	1.15	7	0.0018
extracellular region part	GO:0044421	1099	1.48	8	0.0018
nucleosome	GO:0000786	67	0.09	3	0.0028
extracellular region	GO:0005576	2140	2.89	10	0.0028
protein-DNA complex	GO:0032993	111	0.15	3	0.0159
cytoskeleton	GO:0005856	1790	2.41	8	0.0159
intracellular non-membrane-bounded organelle	GO:0043232	3800	5.13	11	0.0373
Golgi lumen	GO:0005796	77	0.1	2	0.0373
non-membrane-bounded organelle	GO:0043228	3800	5.13	11	0.0373
chromatin	GO:0000785	296	0.4	3	0.0398

Biological Process					
Category	ID	# Genes in category	# Expected	# of genes matched	adjP value
protein-DNA complex assembly	GO:0065004	137	0.21	4	0.0023
nucleosome assembly	GO:0006334	107	0.16	4	0.0023
chromatin assembly	GO:0031497	117	0.18	4	0.0023
nucleosome organization	GO:0034728	128	0.19	4	0.0023
chromatin assembly or disassembly	GO:0006333	138	0.21	4	0.0023
protein-DNA complex subunit organization	GO:0071824	152	0.23	4	0.0024
DNA packaging	GO:0006323	152	0.23	4	0.0024
DNA conformation change	GO:0071103	188	0.28	4	0.0059
peptide cross-linking	GO:0018149	29	0.04	2	0.0237
keratinization	GO:0031424	44	0.07	2	0.0474

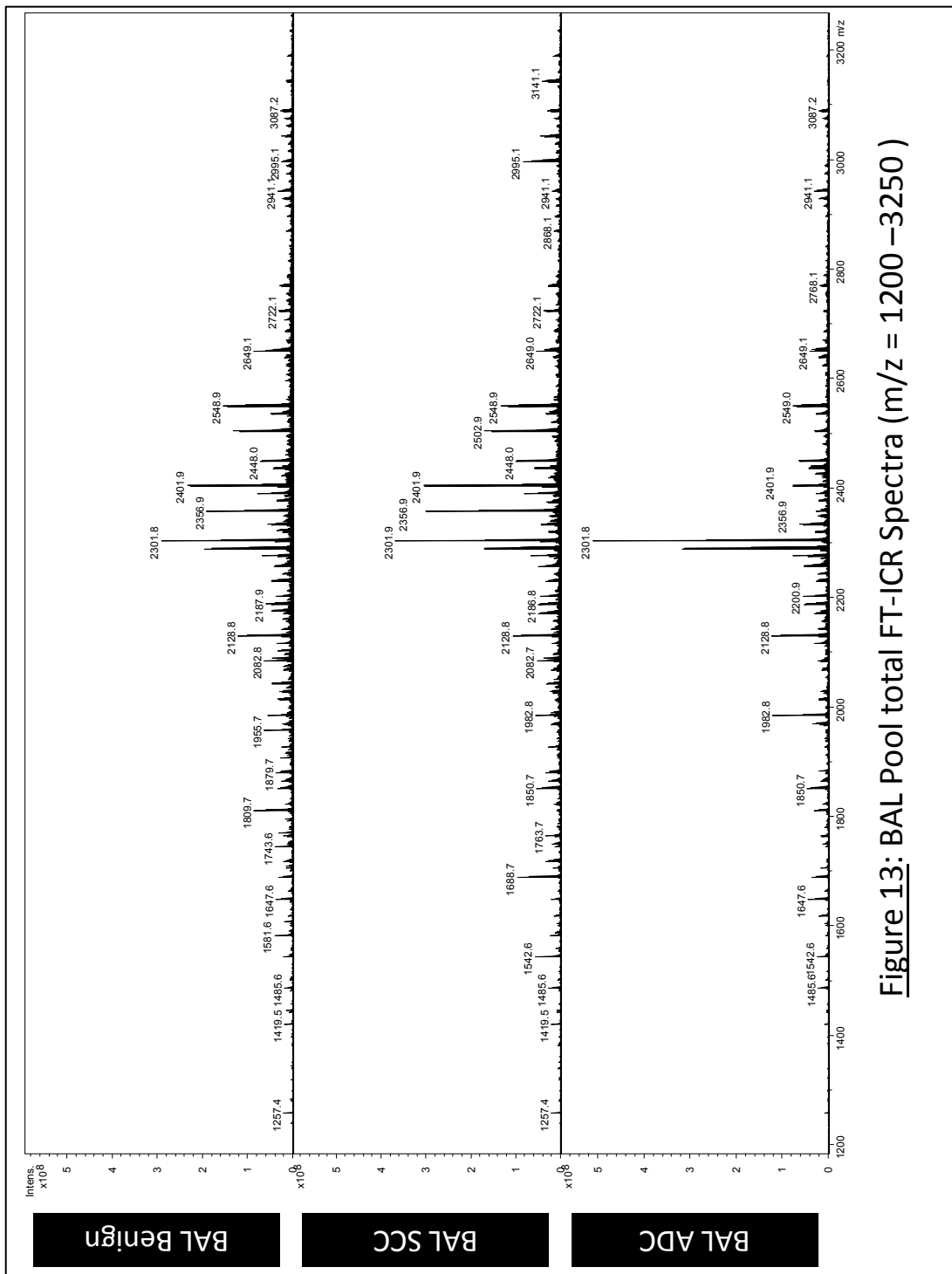


Figure 13: BAL Pool total FT-ICR Spectra ( $m/z = 1200 - 3250$ )

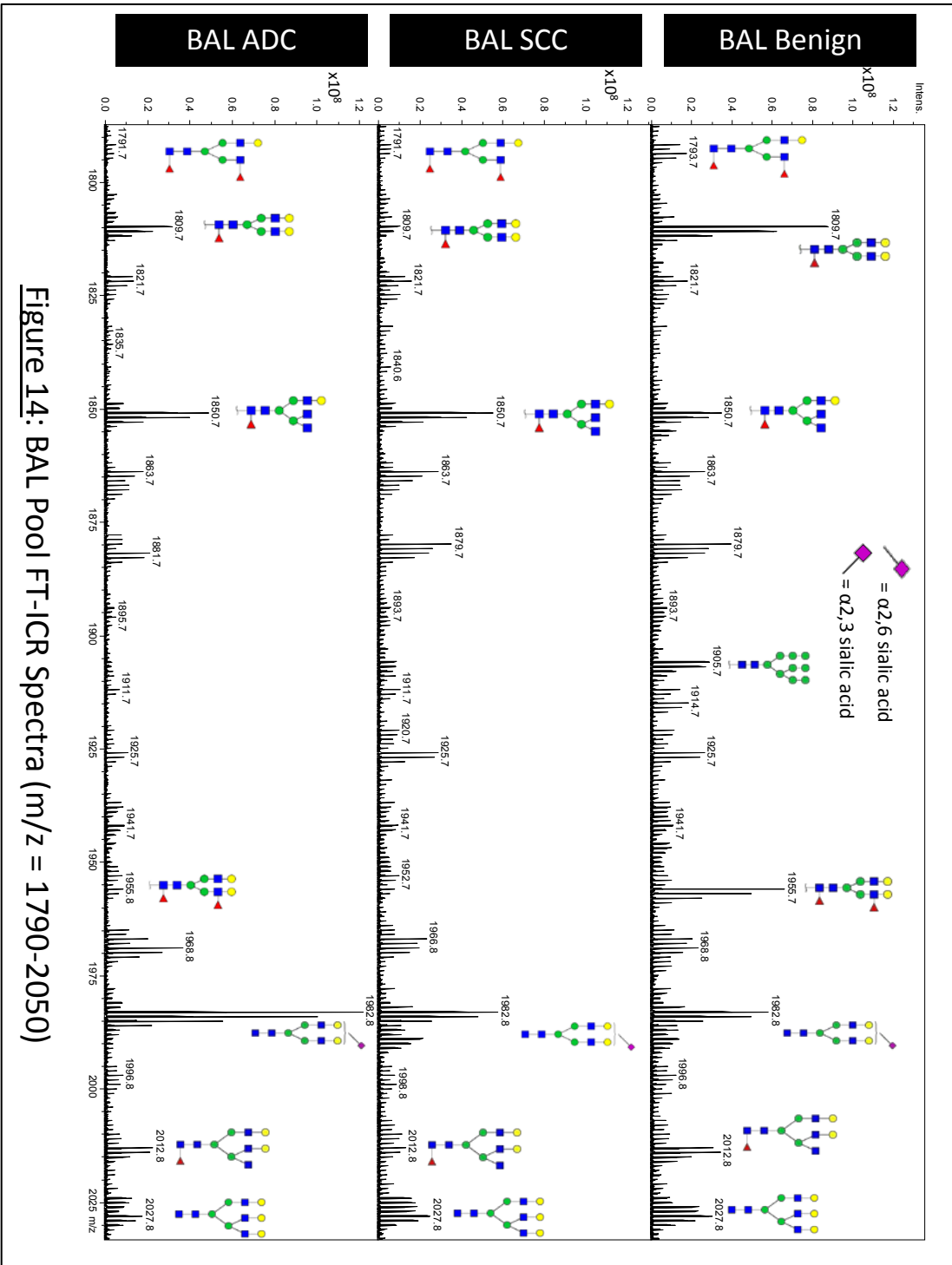


Figure 14: BAL Pool FT-ICR Spectra ( $m/z = 1790-2050$ )

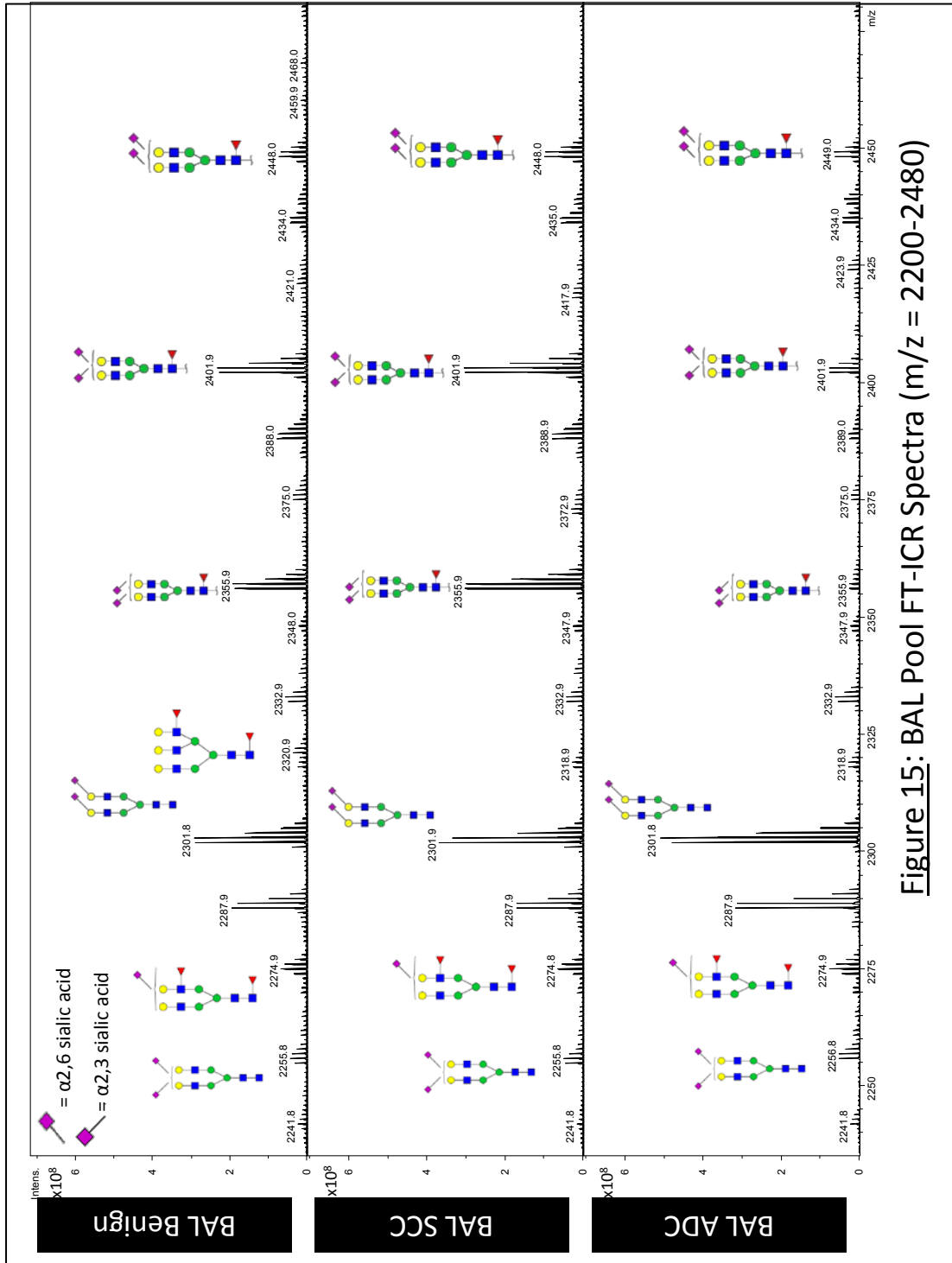


Figure 15: BAL Pool FT-ICR Spectra (m/z = 2200-2480)

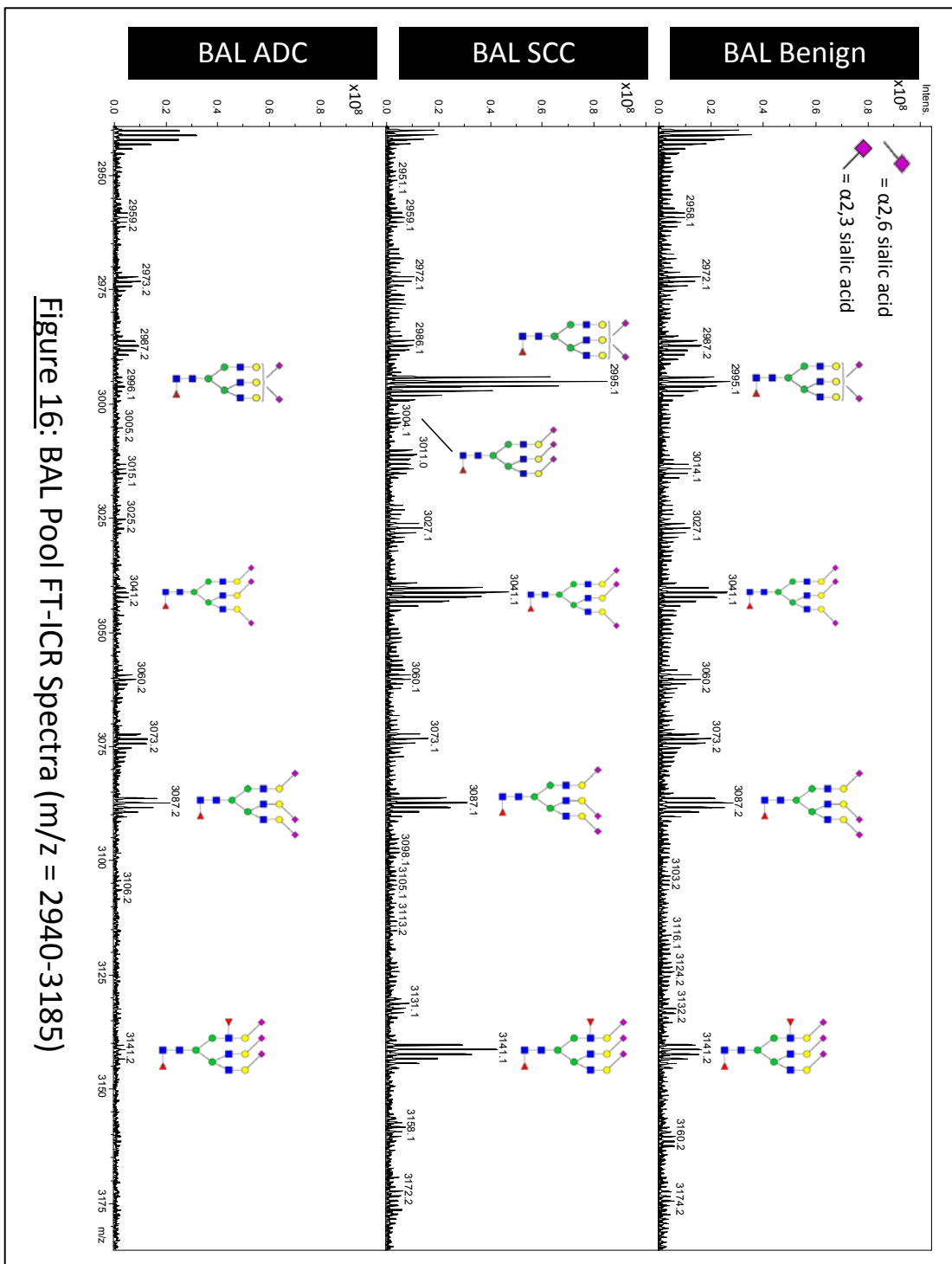
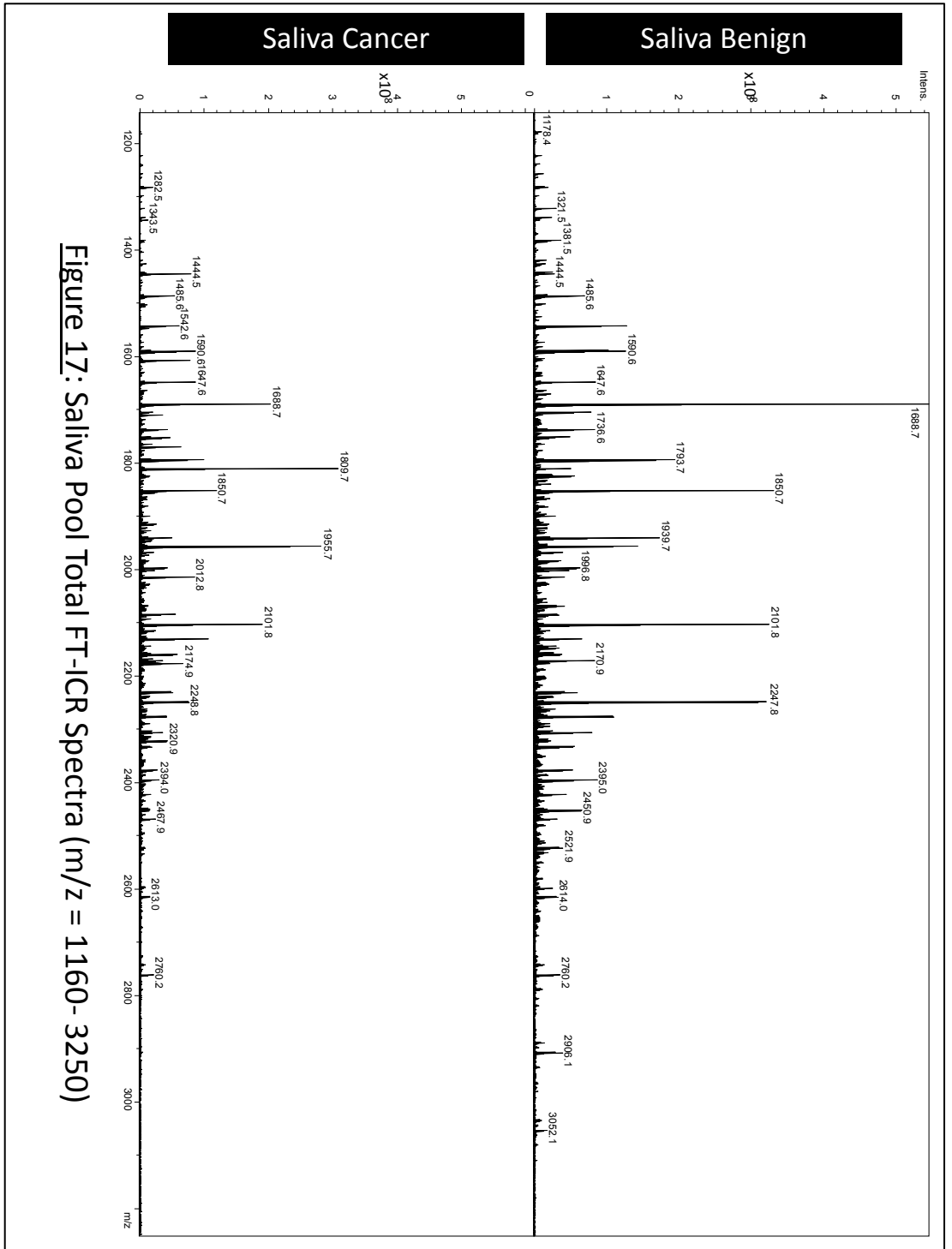


Figure 16: BAL Pool FT-ICR Spectra (m/z = 2940-3185)

Pooled saliva samples were next examined over the mass range of  $m/z = 1160$ - $3250$  (**Figure 17**). Clear differences in the spectra of the benign and cancer were apparent, therefore we assessed which structures were predominant in each group (**Figures 18 - 20**). In the benign pool, the tri-antennary complex glycan Hex3dHex1HexNAc5 ( $m/z = 1688.7$ ) was the most prevalent. This species was followed by a near equal distribution of  $m/z = 1850.7$  and  $2101.8$ , correlating to a core-fucosylated tri-antennary and a tri-fucosylated bi-antennary glycan, respectively (**Figure 18 and 20**). There was also an apparent increase in an  $\alpha$ 2-6-mono-sialylated tri-antennary glycan ( $m/z = 2169.8$ ) and a decrease in a di-fucosylated tri-antennary complex species ( $m/z = 2320.9$ ) in the benign pool relative to the cancer (**Figure 20**). The cancer pool presented with a slightly higher mass tri-antennary glycan, Hex4dHex1HexNAc4 ( $m/z = 1809.7$ ) dominating peak height. In addition, the increased relative abundance of bi-antennary Hex5dHex2HexNAc4 ( $m/z = 1955.7$ ) and tri-antennary Hex6dHex1HexNAc5 ( $m/z = 2174.9$ ) in the cancer samples (**Figures 19 and 20**).



**Figure 17: Saliva Pool Total FT-ICR Spectra (m/z = 1160- 3250)**

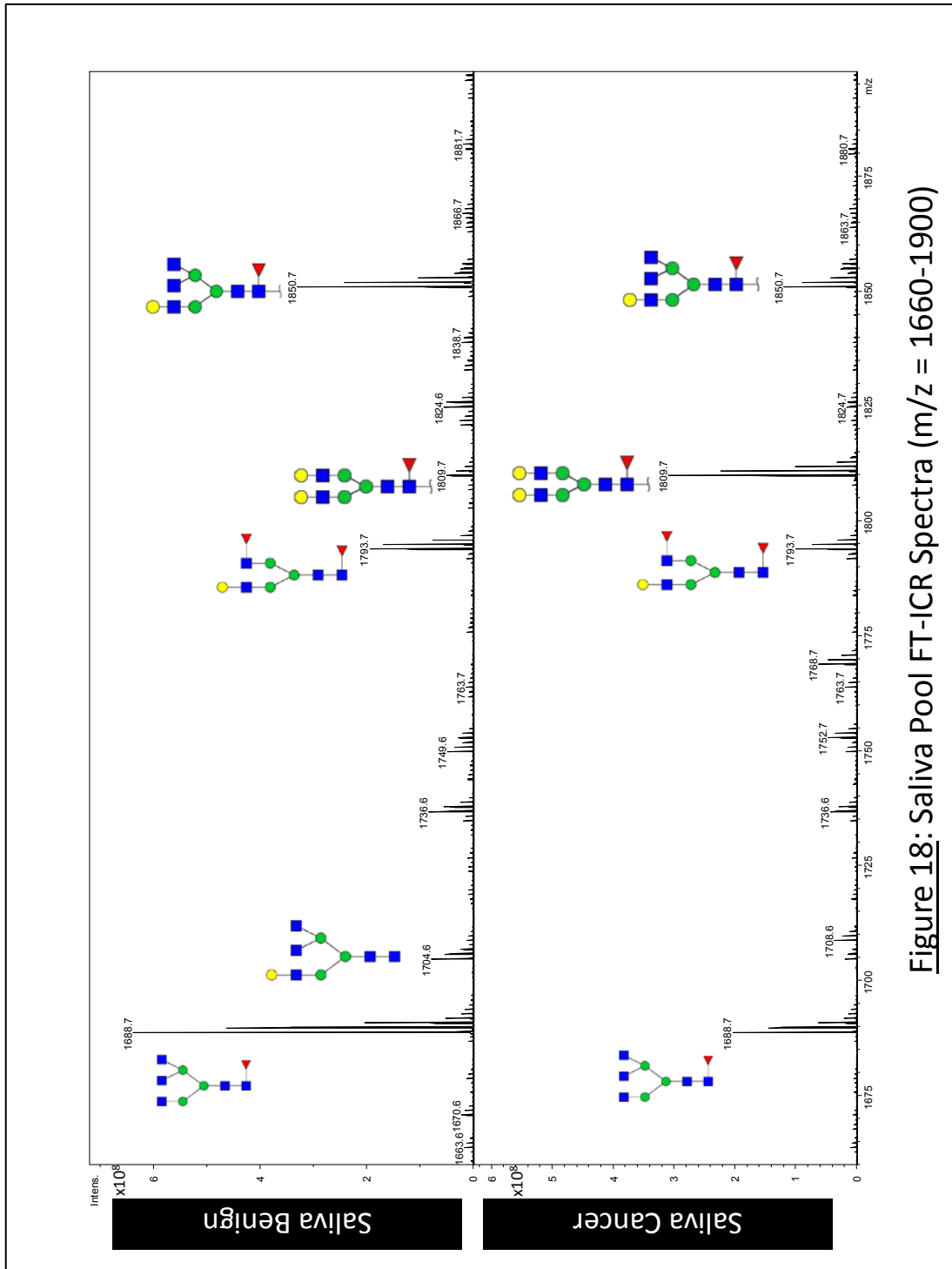


Figure 18: Saliva Pool FT-ICR Spectra (m/z = 1660-1900)



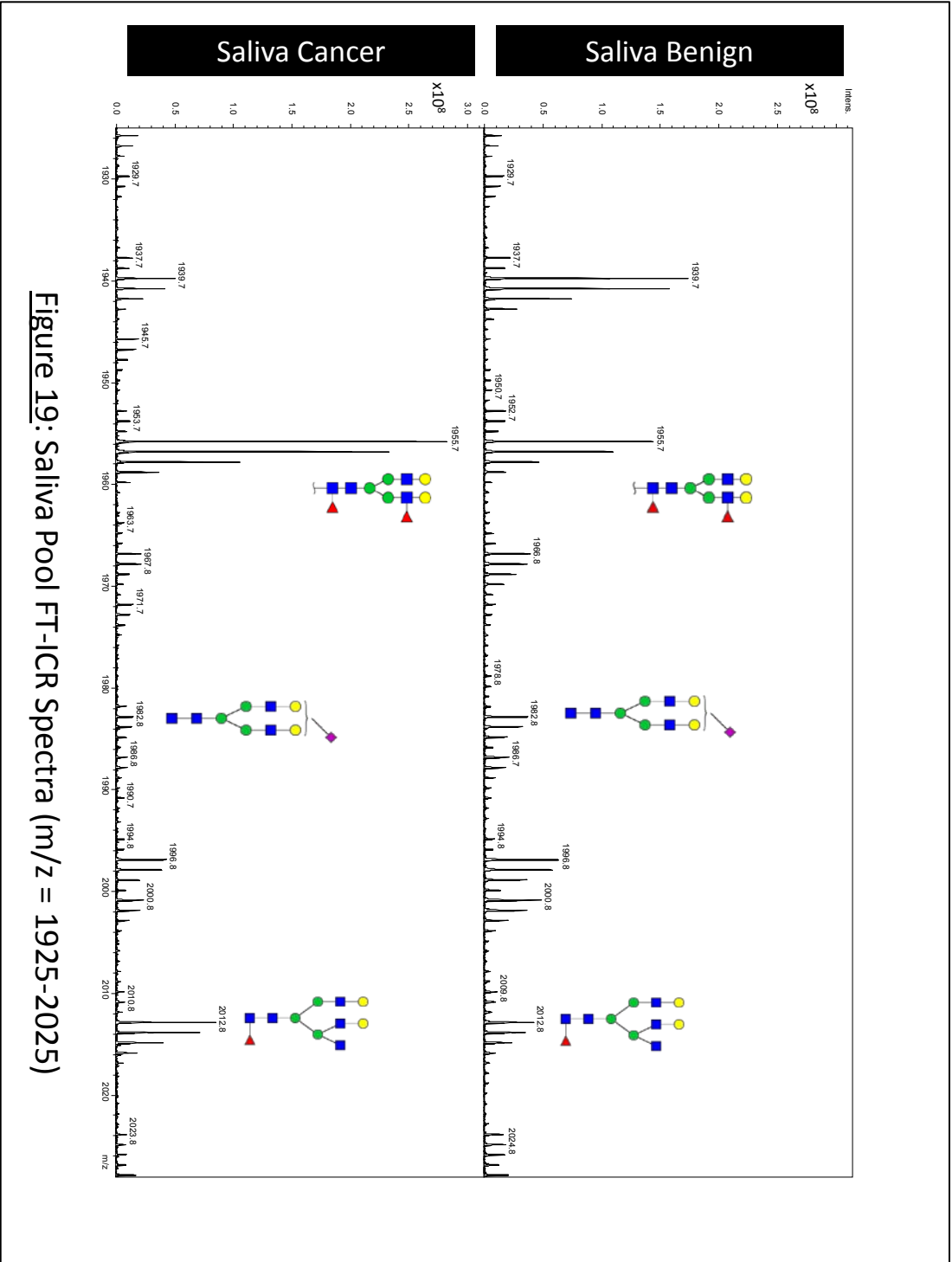


Figure 19: Saliva Pool FT-ICR Spectra (m/z = 1925-2025)

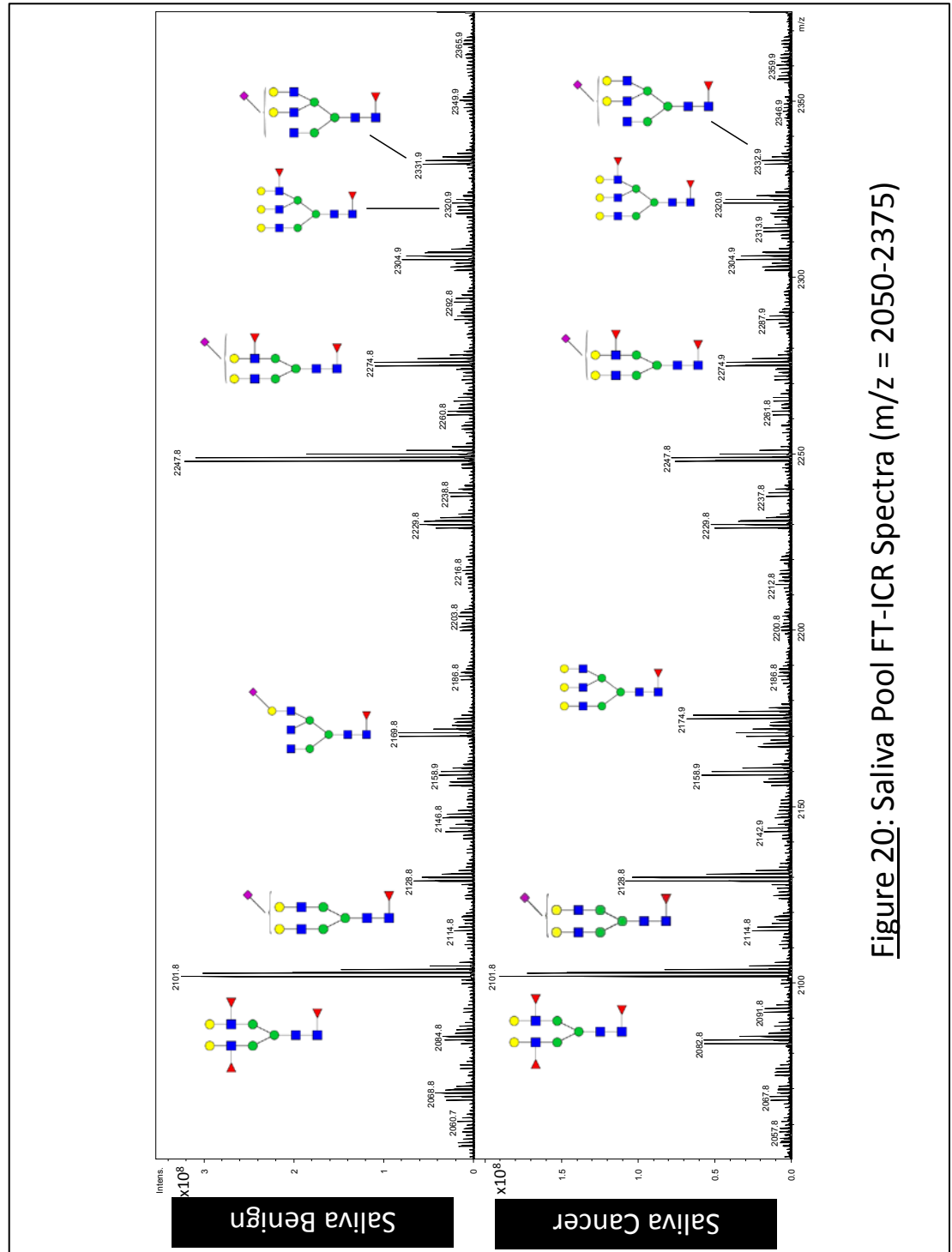
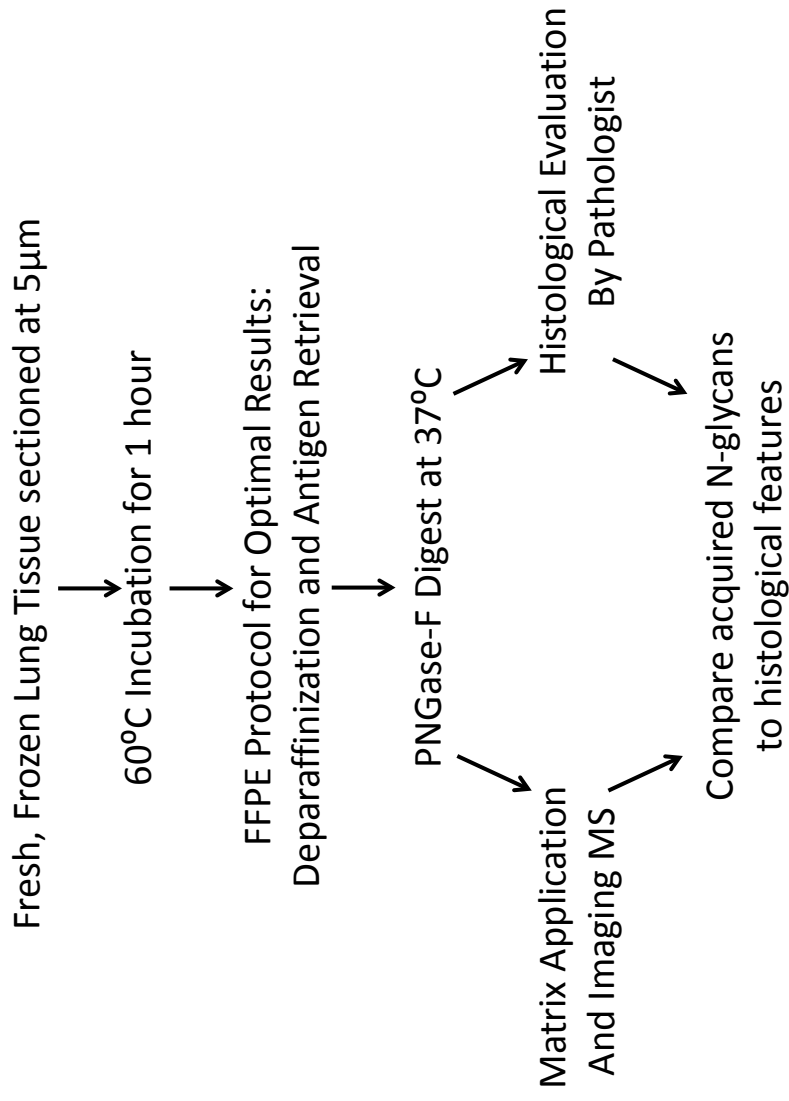


Figure 20: Saliva Pool FT-ICR Spectra (m/z = 2050-2375)

### 4.3.3 N-linked Glycan Tissue Imaging of NSCLC Tissue

Human lung tissue was sectioned and processed for on-slide PNGase-F digest, matrix application, and MALDI-FT-ICR acquisition of released N-glycans (**Figure 21**). These included tumor and adjacent normal tissue from one adenocarcinoma and two squamous cell carcinomas. A duplicate slide of each was stained with hematoxylin and eosin and histological features were confirmed by a pathologist (**Figure 22**). The distribution of individual peaks was compared against defined regions to determine how their expression correlated to histology (**Figures 23a and 23b**). From this effort, we identified three glycans whose presence was increased in the tumor tissues (both ADC and SCC) relative to normal adjacent ( $m/z = 1419.5, 1743.6, \text{ and } 2174.8$ ). These structures correspond to Hex6HexNAc2, Hex8HexNAc2, and Hex6dHex1HexNAc5, respectively. In addition, three other species appeared to be up-regulated exclusively in the defined cancer region of ADC ( $m/z = 1955.7, 2320.8, \text{ and } 2686.0$ ). These glycans correlated to a set of di-fucosylated structures including Hex5dHex2HexNAc4, Hex6dHex2HexNAc5, and Hex7dHex2HexNAc6. There was also a single spectra, correlating to Hex3dHex1HexNAc4, which was selectively dominant in SCC tumor tissues ( $m/z = 1485.5$ ). In addition, we identified several N-glycans whose expression did not appear to differentiate tumor from normal lung tissue (**Figure 23a**;  $m/z = 1257.4$ ).



**Figure 21:** Workflow for Lung Tissue Imaging of N-glycans  
Adapted from Powers et al. PLOS ONE. Vol.9 (9). Sept. 2014.



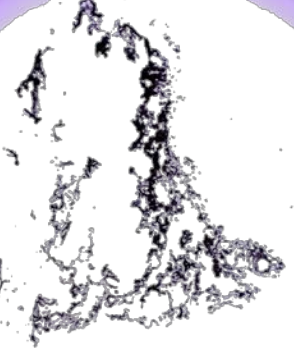
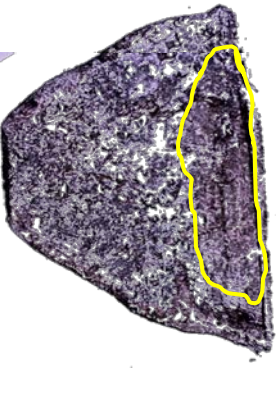
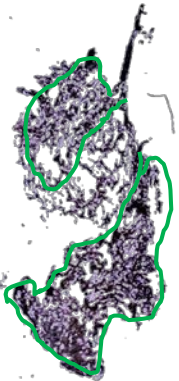
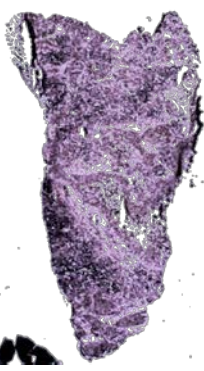
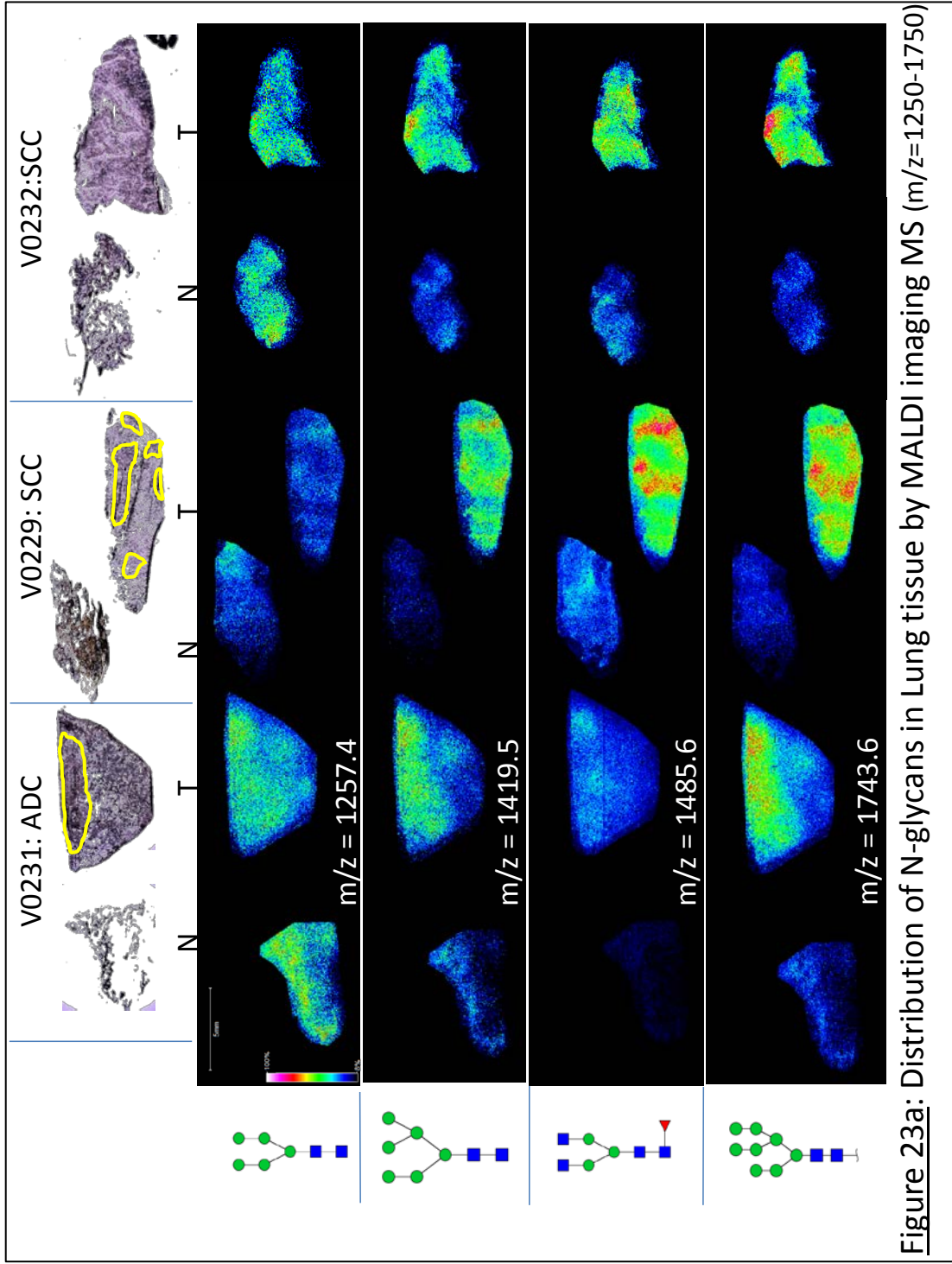
<p style="text-align: center;">Normal</p> <p>V0229: SCC</p>  <p style="text-align: center;">Inflammation throughout</p>	<p style="text-align: center;">Tumor</p>  <p style="text-align: center;">Tumor (yellow)</p>
<p>V0231: ADC</p> 	 <p style="text-align: center;">Tumor (yellow). Remainder is with inflammation</p>
<p>V0232: SCC</p>  <p style="text-align: center;">Inflammation (green)</p>	 <p style="text-align: center;">Cancer cell nests mixed with stroma (50:50)</p>

Figure 22: Lung Tissue H&E with Pathology Notes



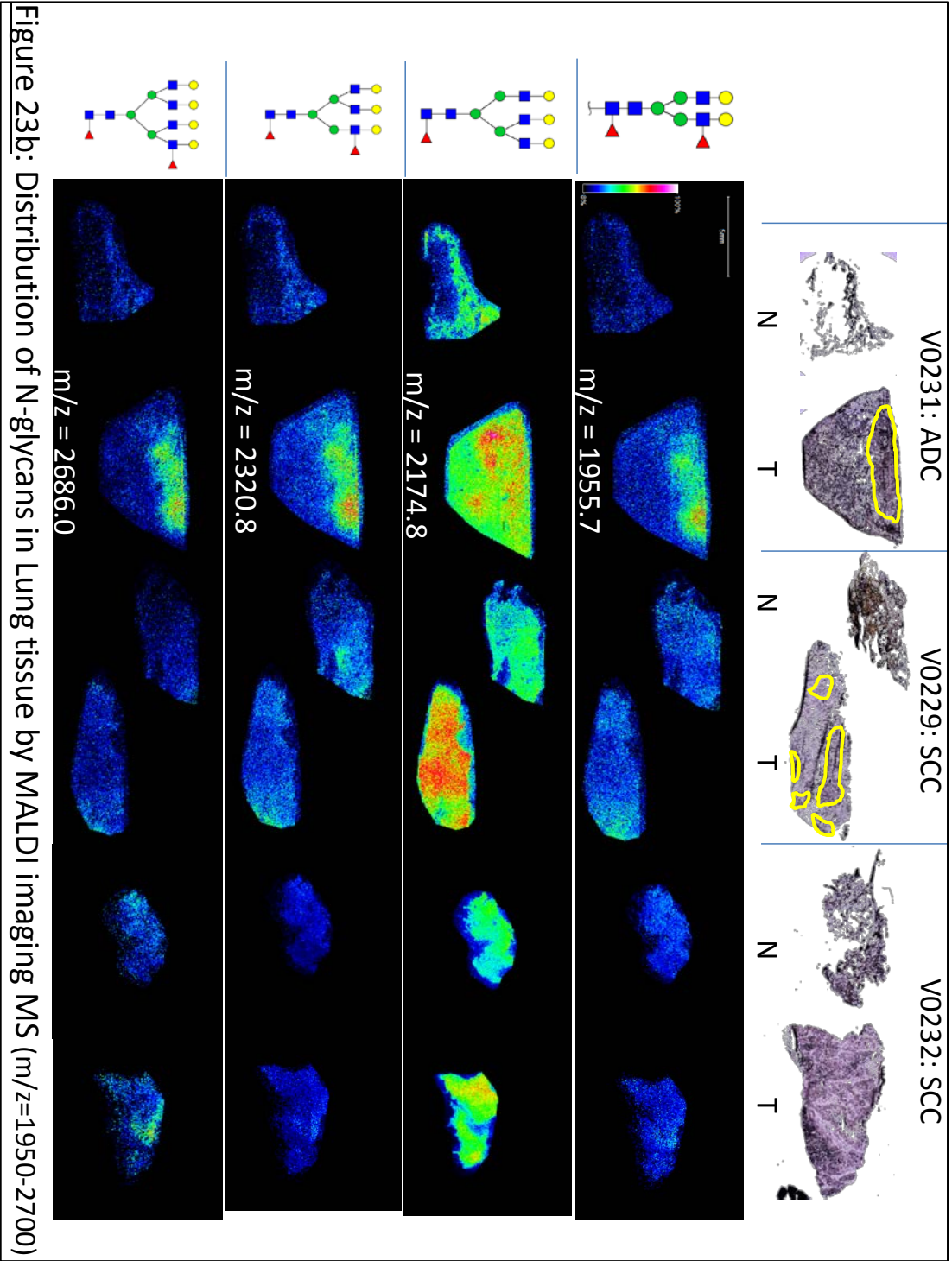


Figure 23b: Distribution of N-glycans in Lung tissue by MALDI imaging MS ( $m/z=1950-2700$ )

## 4.4 Discussion

Here the application of recently-developed methods for N-linked glycomics on pathologically-graded, clinically-relevant NSCLC samples was assessed. A pre-analytical sialic acid linkage derivatization approach was applied to biofluids, including BAL and saliva, and acquired mass spectra using a high resolution accurate mass instrument. We also performed an on-tissue PNGase-F digest and imaging mass spectrometry to spatially resolve glycans on tumor and adjacent normal tissues. The rationale for these experiments was seeded both in the methodology employed and the types of samples analyzed; to gauge their collective utility for future biomarker discovery efforts.

The proteome of human BAL pools (benign and cancer) was first evaluated to catalogue the components detectable in each diagnostic group by LC-MS/MS. This substrate has previously demonstrated relevance as a tool for diagnostic/prognostic evaluations of various pathologies, however explorations of its proteomic content are still rather new(15,233,234,241). We discovered over 300 proteins across the 10 samples analyzed and applied label-free quantification as a means initially assess relative abundance of each. General trends were observed in protein abundance based on the application of label-free quantification which would likely be of value exploring using isotopic labeling techniques such as, iTRAQ, or TMT methods. In addition, based on the high nSC values obtained for albumin, IgG, and transferrin-related proteins, the application of up-front depletion/enrichment strategies would likely increase the depth of identifications for sub-proteomic investigations.

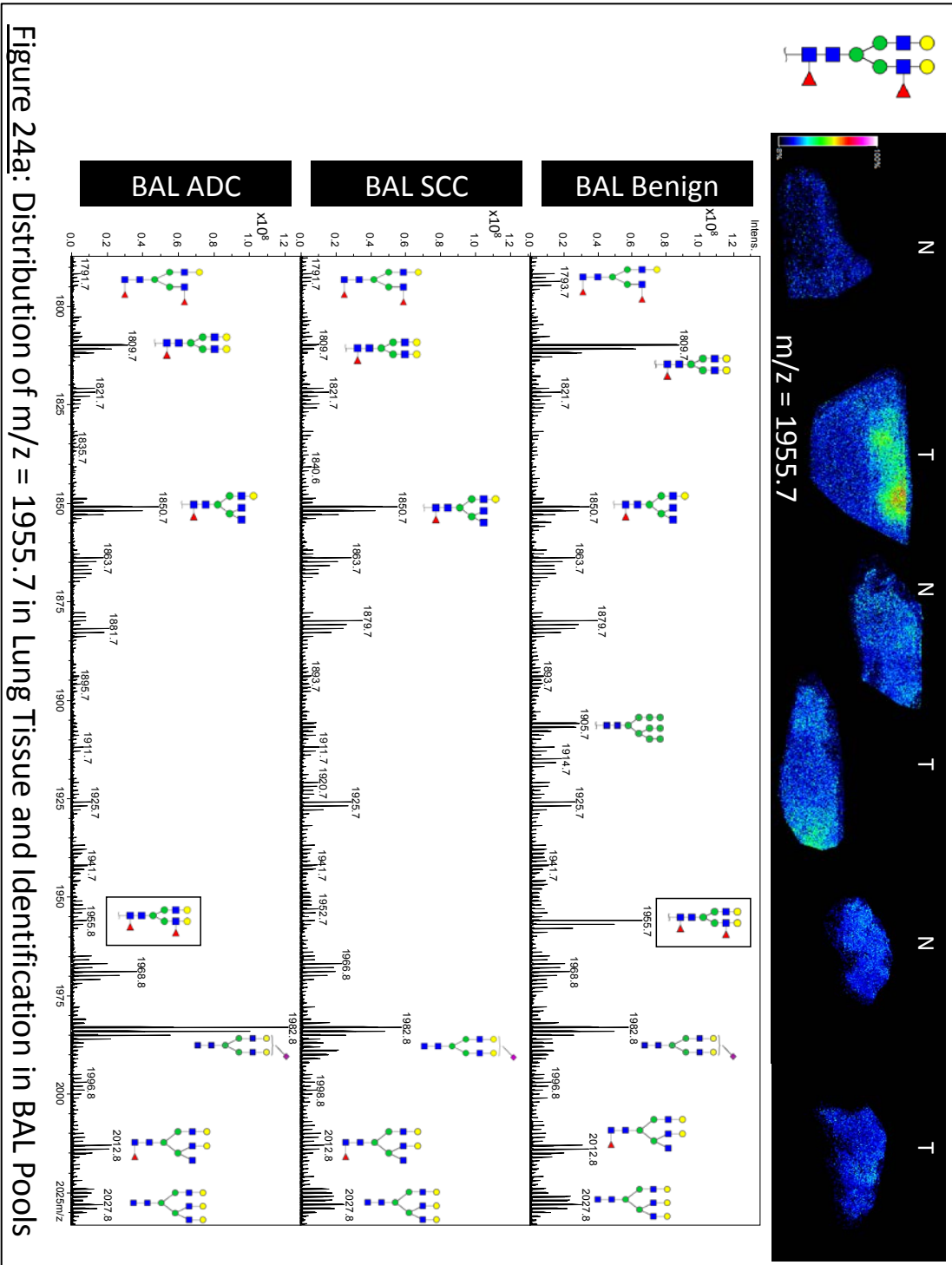
The N-linked glycome of BAL fluid was analyzed using sialic acid linkage specific esterification and MALDI-FT-ICR. This was the first assessment of N-linked

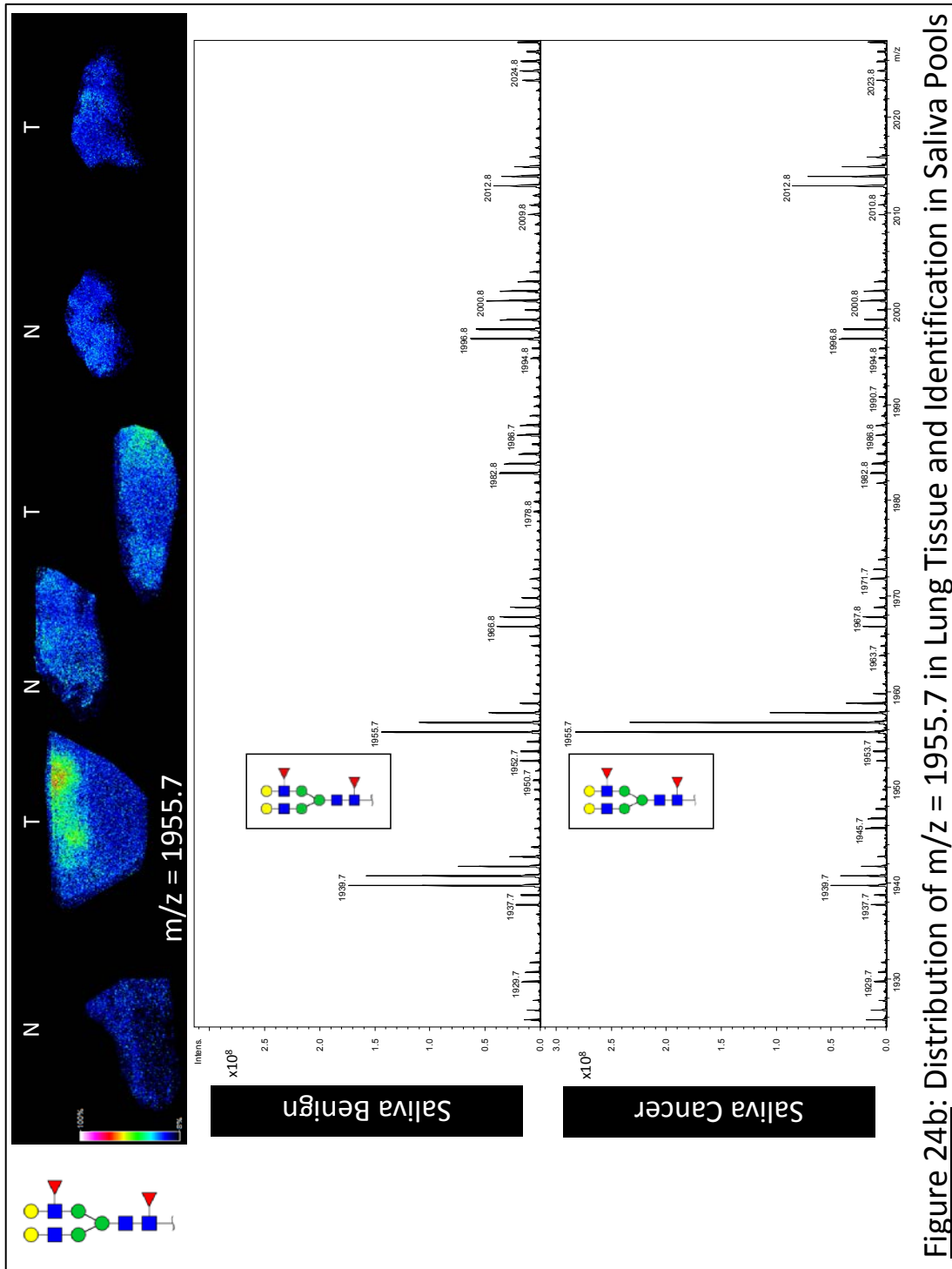


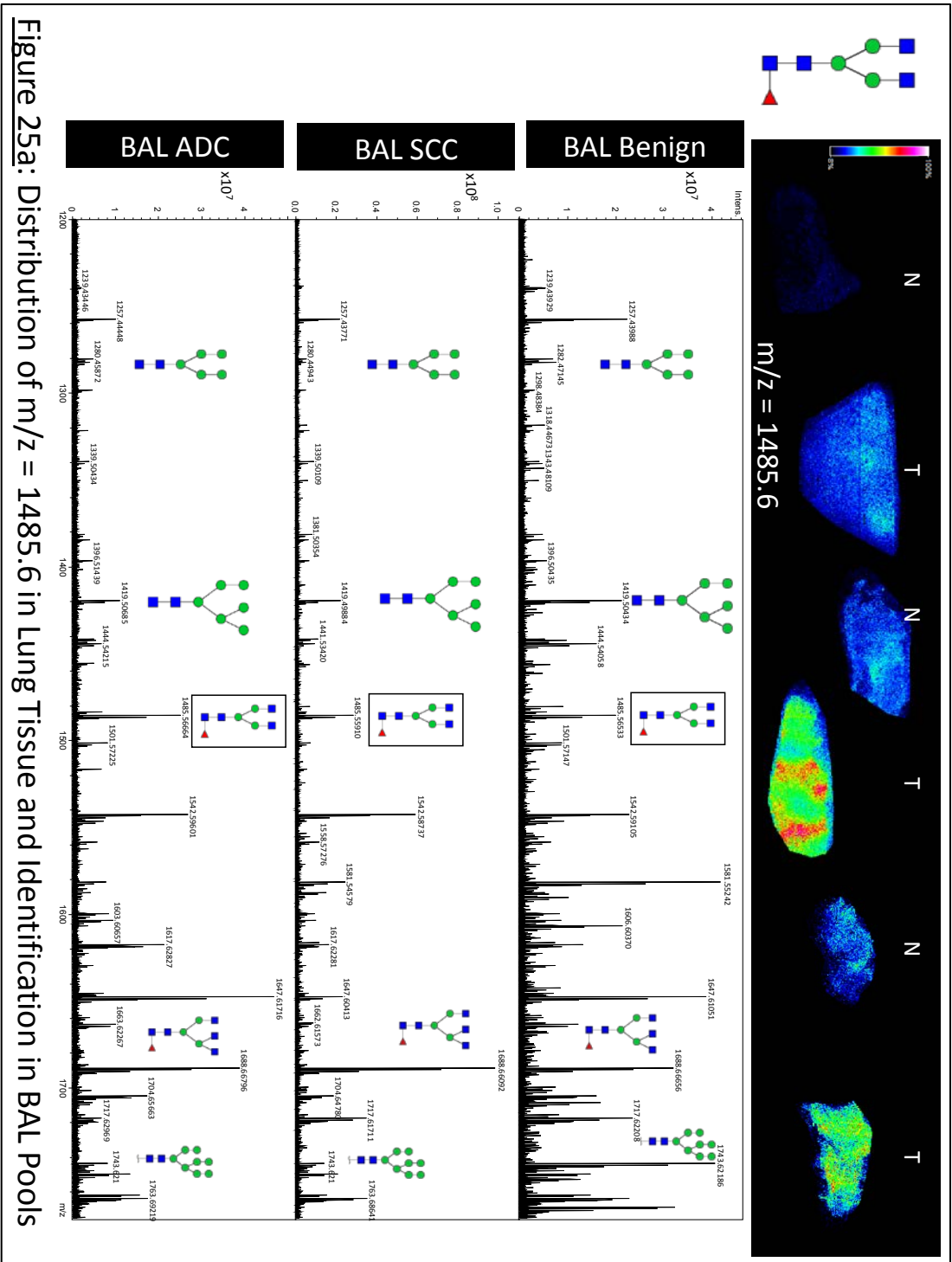
glycoforms derived from this proximal fluid that we know of. In addition, we profiled structures present in matched saliva pooled samples to increase the breadth of investigation and discern the utility of these substrates for glycomics-based biomarker discovery efforts in NSCLC. An obvious difference in the distribution of peaks associated with N-linked glycan structures was apparent between BAL pools. The glycan structural database from the consortium for functional glycomics was utilized to make identifications based on FT-ICR accurate mass detection. Several bi- and tri-antennary complex structures, including core-fucosylated and mono-, di-, and tri-sialylated forms, were shared between pools. Based on relative peak intensity, we discovered the predominant species present in each group was a di- $\alpha$ 2-6-sialylated bi-antennary complex glycan (**Figure 13**;  $m/z = 2301.8$ ). Three glycans, including an oligomannose structure, appeared to be uniquely up-regulated within the benign pool (**Figure 14**). As well, there was apparent down regulation of two structures (**Figure 15**;  $m/z = 2355.9$  and  $2401.9$ ) in the ADC pool relative to benign and SCC. Furthermore, the application of linkage specific sialic acid esterification resolved the presence of di- and tri- sialylated complex glycans, displaying multiple anomeric configurations, in the higher mass ranges of these samples (**Figure 16**). For saliva pools, there was a generalized decrease in the number of spectra correlating to sialylated glycan moieties relative to the BALs. Structures identified in saliva-based benign and cancer groups were predominantly lower mass bi- and tri- antennary complex glycans. Relative peak height suggested that the predominant species in the benign pool was by far a core fucosylated tri-antennary moiety, while the cancer group appeared to have a more even distribution of 3 main structures (**Figure 17**). Interestingly, 11 shared glycan structures were identified in both the saliva and BAL

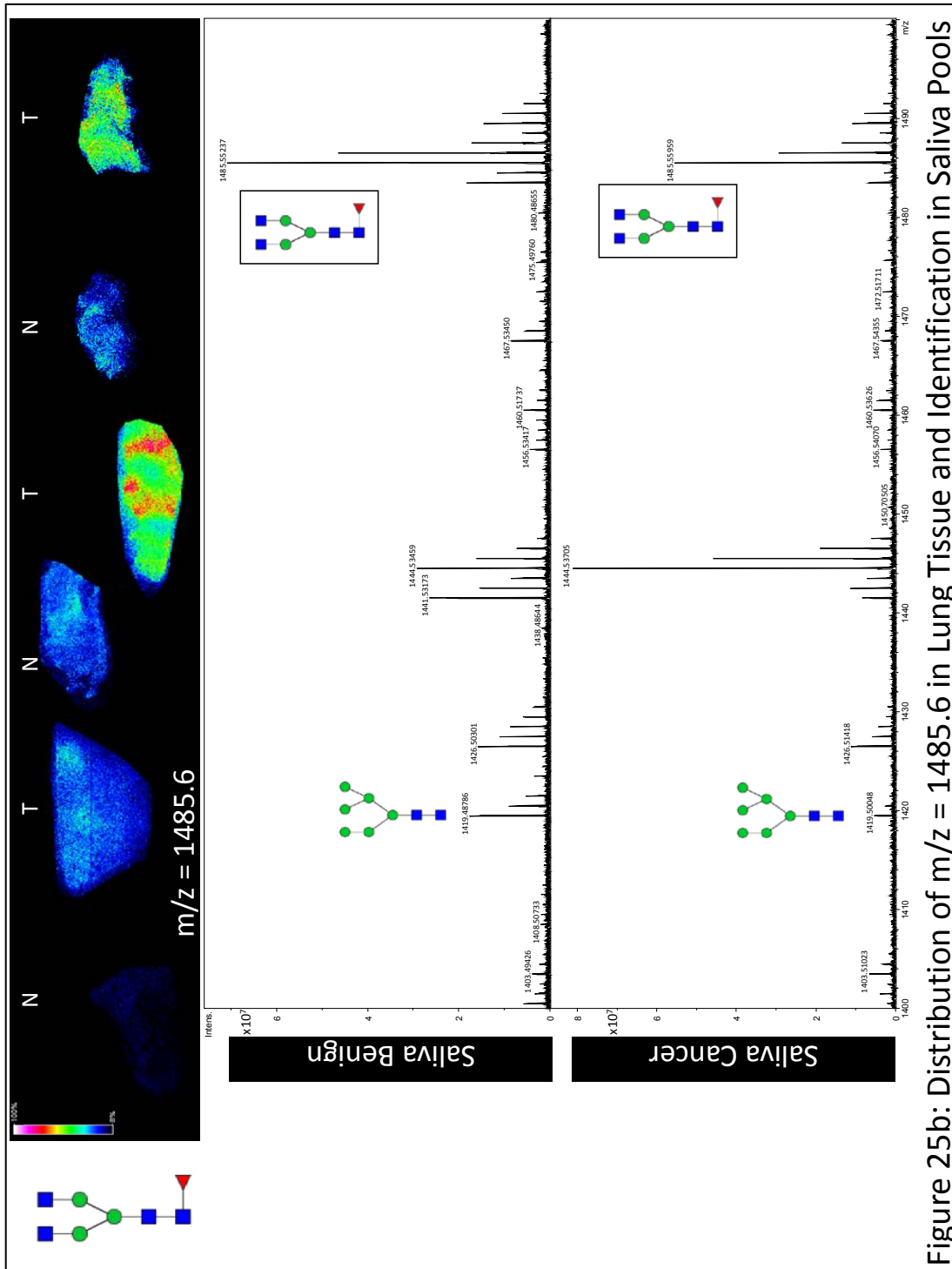
pools. These included four of the most abundant peaks identified in saliva representing three fucosylated bi-antennary and one fucosylated tri-antennary complex glycans (**Figures 18 and 19**;  $m/z = 1793.7, 1809.7, 1850.7, \text{ and } 1955.7$ ).

Intact lung tissue MALDI-FT-ICR imaging mass spectrometry was used to spatially profile the distribution of native N-glycans in tumor and normal-adjacent tissues. Eight oligomannose structures were presented as a demonstration of how this data could be useful for defining specific features of histology. In addition, we cross-referenced several peaks from the imaging data which had previously been identified in the proximal fluid studies. These included species whose localization appeared to correlate to increased expression in ADC only (**Figure 24a and 24b**;  $m/z = 1955.7$ ), SCC only (**Figure 25a and 25b**;  $m/z = 1485.6$ ), or within both cancer groups cancer (**Figure 26a and 26b**;  $m/z = 1419.5$ ). Conversely, we assessed potentially interesting spectra from the proximal fluids in the imaging data. For example, based on visual assessment of peak height, the core-fucosylated tri-antennary glycan ( $m/z = 1688.7$ ) was the prevalent species in the benign saliva pool (**Figure 17**). However, the spatial distribution of this species, and peak height in the BAL pools, did not correlate with this observation (**Figure 27**). In a similar manner, the high-mannose glycan ( $m/z = 1905.7$ ) appeared to be more abundant in the benign BAL pools relative to SCC and ADC; yet imaging data suggested a higher association with tumor tissues (**Figure 28**).









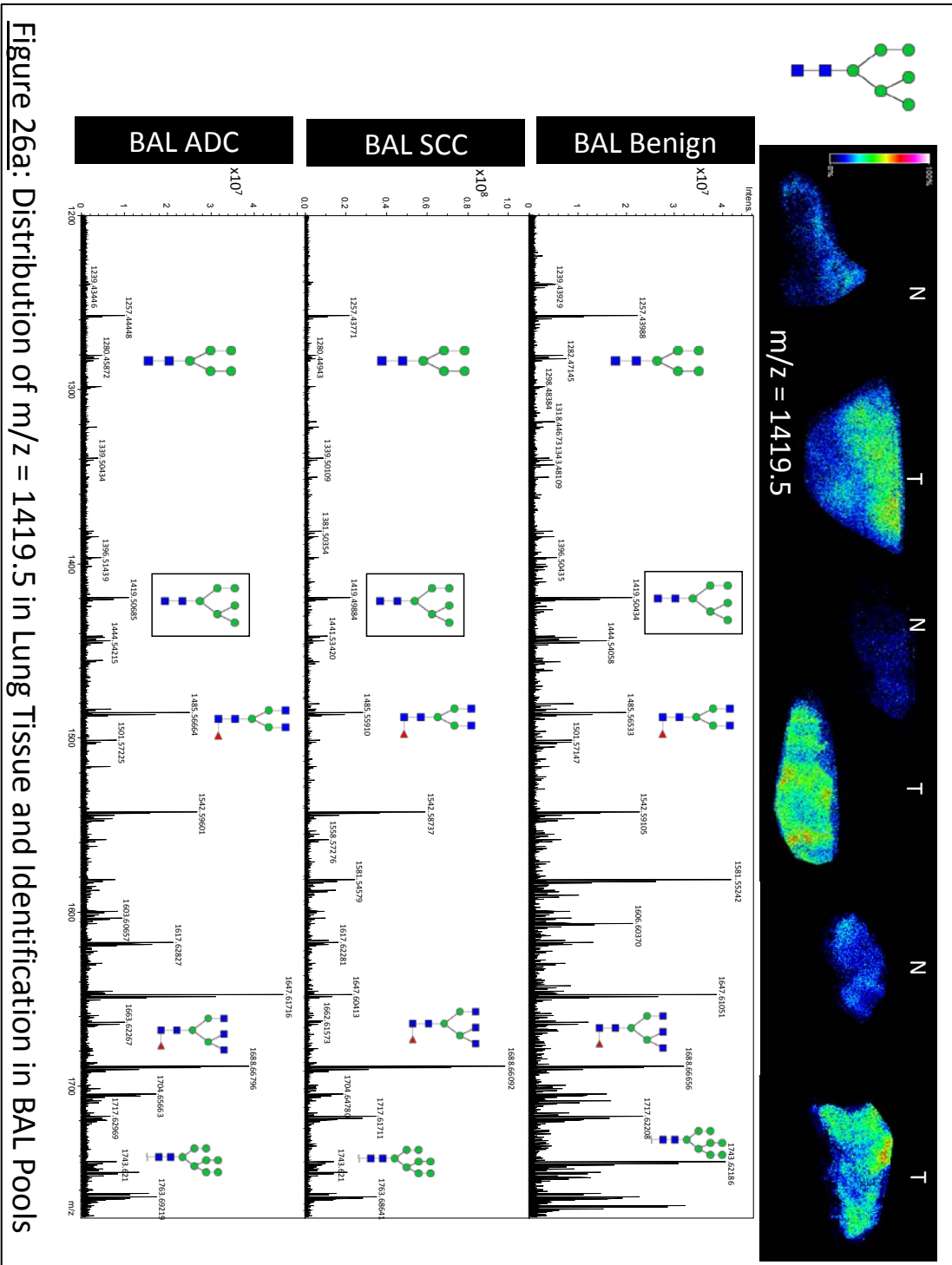
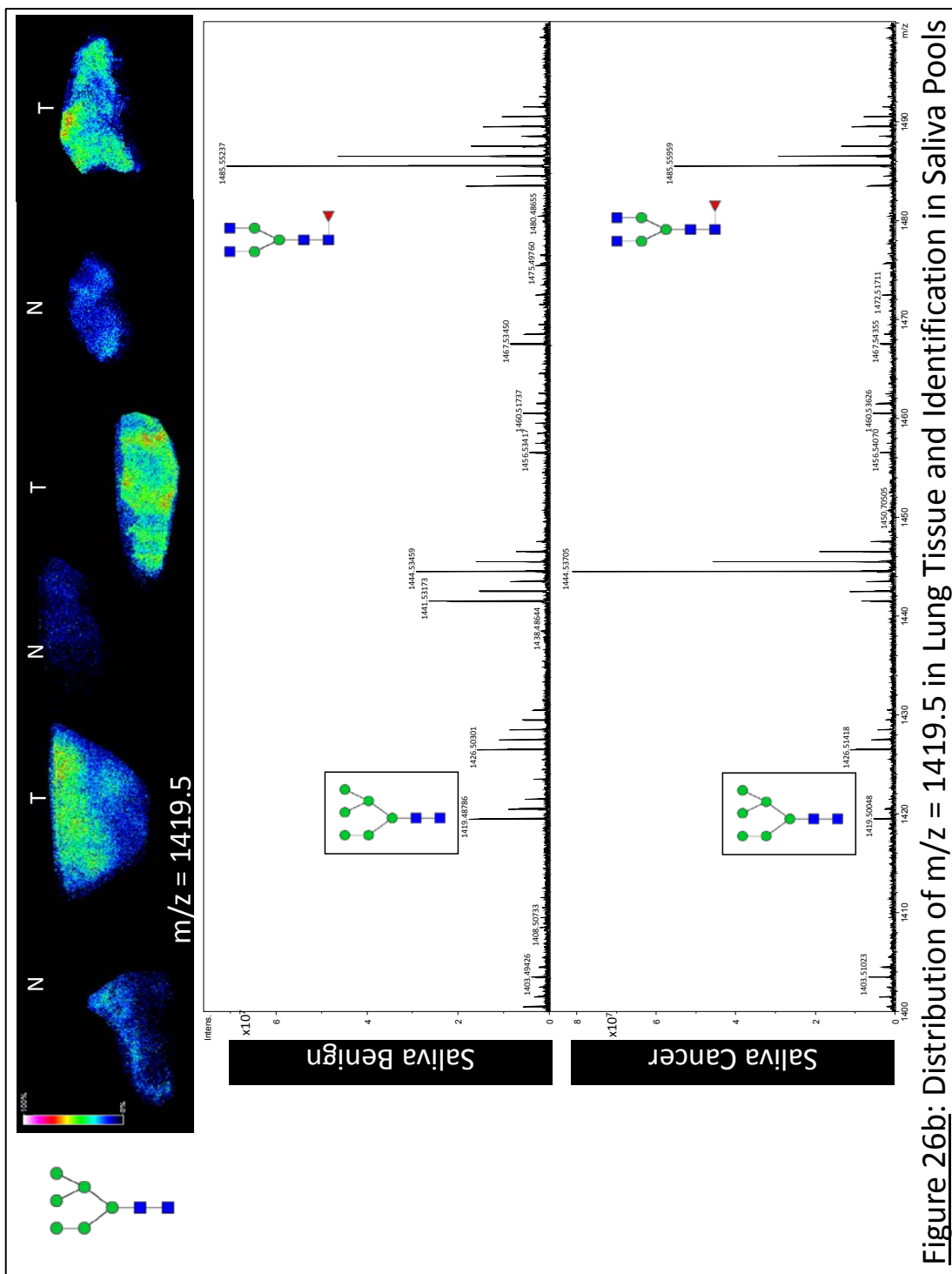
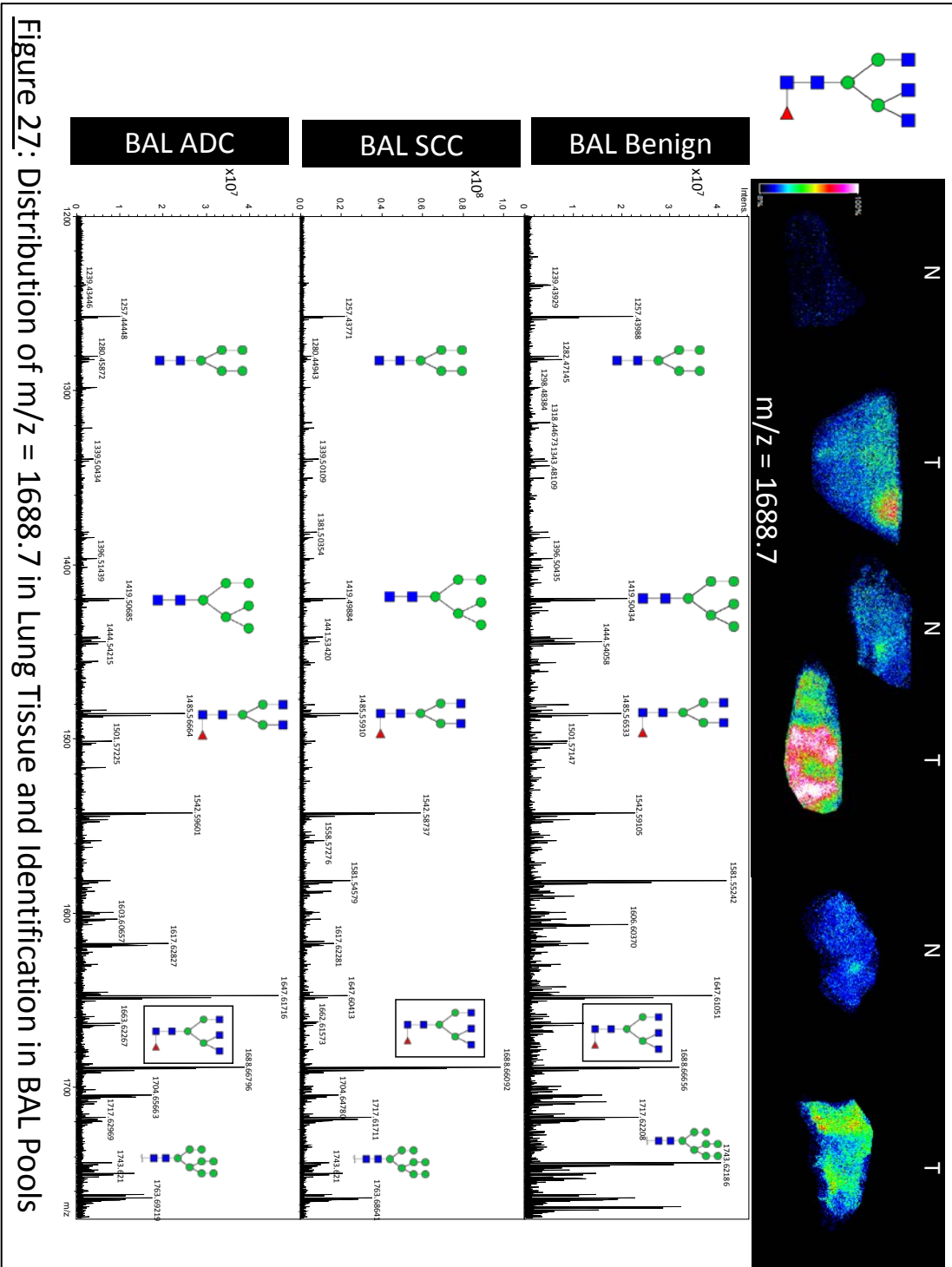
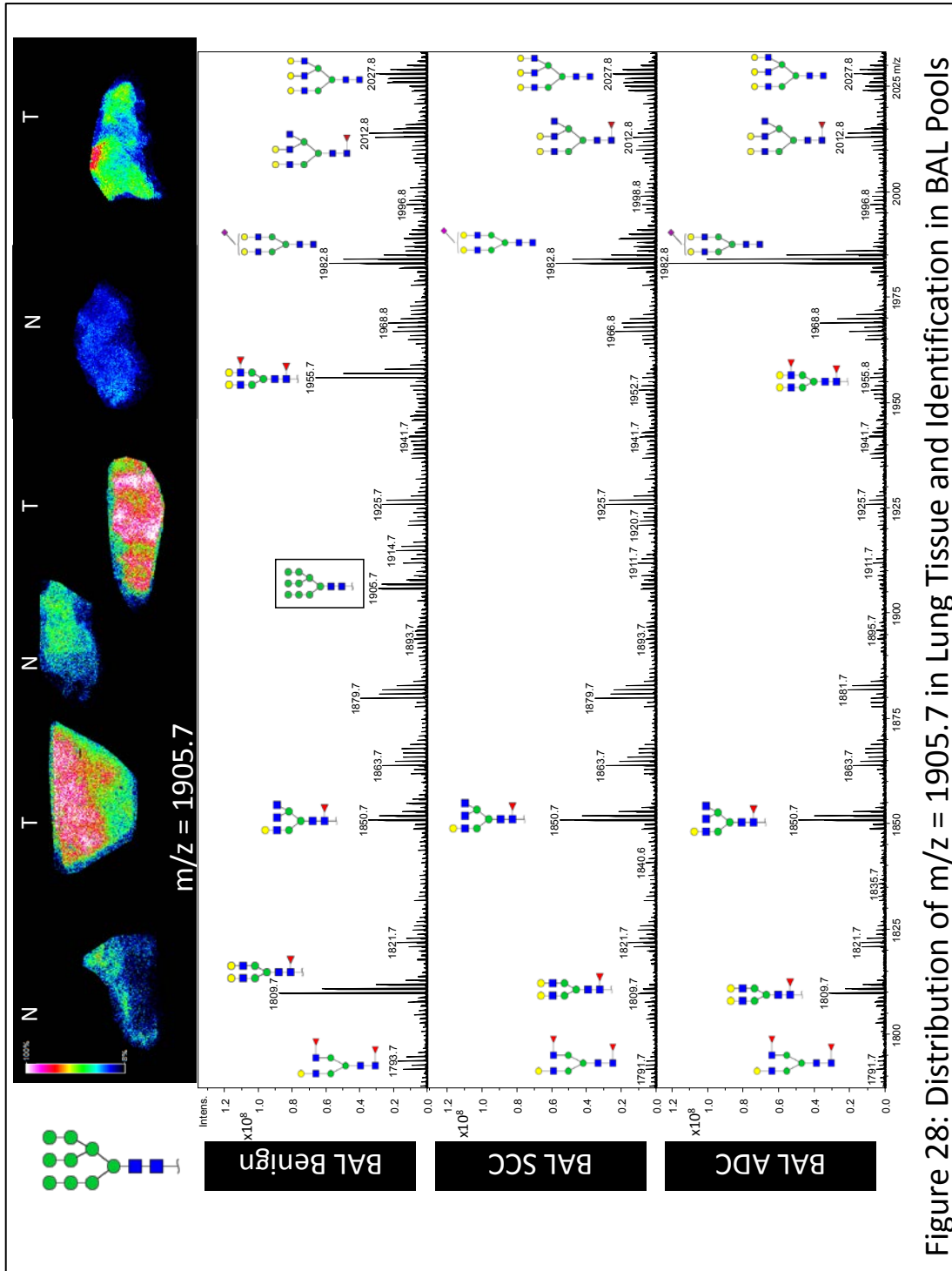


Figure 26a: Distribution of  $m/z = 1419.5$  in Lung Tissue and Identification in BAL Pools









Obvious caveats limit the ability to assess trends in glycan expression between substrates as presented. These include the use of pooled proximal fluid samples and the low number of tissues used for imaging. In addition, statistical analysis for evaluating differences in relative peak height was not performed and reproducibility has yet to be investigated. Furthermore, the release of glycans from dominant endogenous glycoproteins in these substrates may overpower the detection of lower abundance species. Another limiting factor related to these investigations was the use of linkage-specific sialic acid derivatization for proximal fluids only. This approach was adopted to both protect sialic-acid containing glycans during ionization and discern their anomeric configurations. However, an equivalent method for on-tissue derivatization has yet to be developed; precluding extensive detection of sialylated species for on-tissue experiments. As an alternative, increased identification of these glycans might be achieved by imaging mass spectrometry carried out in negative ionization mode(242).

Bronchoalveolar lavage fluid and saliva are appropriate substrates for the discovery of disease-specific glycosylation patterns in NSCLC. These fluids are abundant sources of secreted/shed glycoproteins and are frequently collected during routine diagnostic or follow-up procedures. Saliva may be acquired regularly in copious volumes with minimum patient risk, while BAL fluids are typically available from patients presenting with a respiratory mass following bronchoscopy. Although BAL has been probed for diagnostic and prognostic utility based on cytological and (more recently) proteomic content, investigations of N-linked glycoforms in this fluid are novel(233,234,243,244). Similarly, efforts to utilize saliva for biomarker discovery have demonstrated promising results in lung, head and neck, and breast cancer(245–249).

Saliva is known to be an rich source of carbohydrate molecules stemming from mucins and high molecular weight glycoproteins, however explorations into the types of glycan structures it carries, and how they relate to specific pathologies, are limited(188,247,249,250). In addition, the presented methods could also be applied to proximal fluids such as exhaled breath condensate or pleural fluid. These materials contain proteomic content derived from the respiratory epithelium and early studies have demonstrated their feasibility for use in biomarker evaluations(251,252). While all of these fluids may be potent reservoirs of glycan molecules related to disease, consideration must be given to collection and processing procedures in order to reliably assess reproducibility of findings. This potential limitation will be addressed in our on-going studies by continuing to use bio-repository maintained samples with well-documented collection/processing procedures. The use of this sample resource will also provide us with follow-up patient data which will likely be of value for correlating glycan structures to specific outcomes such as disease recurrence and mortality rates.

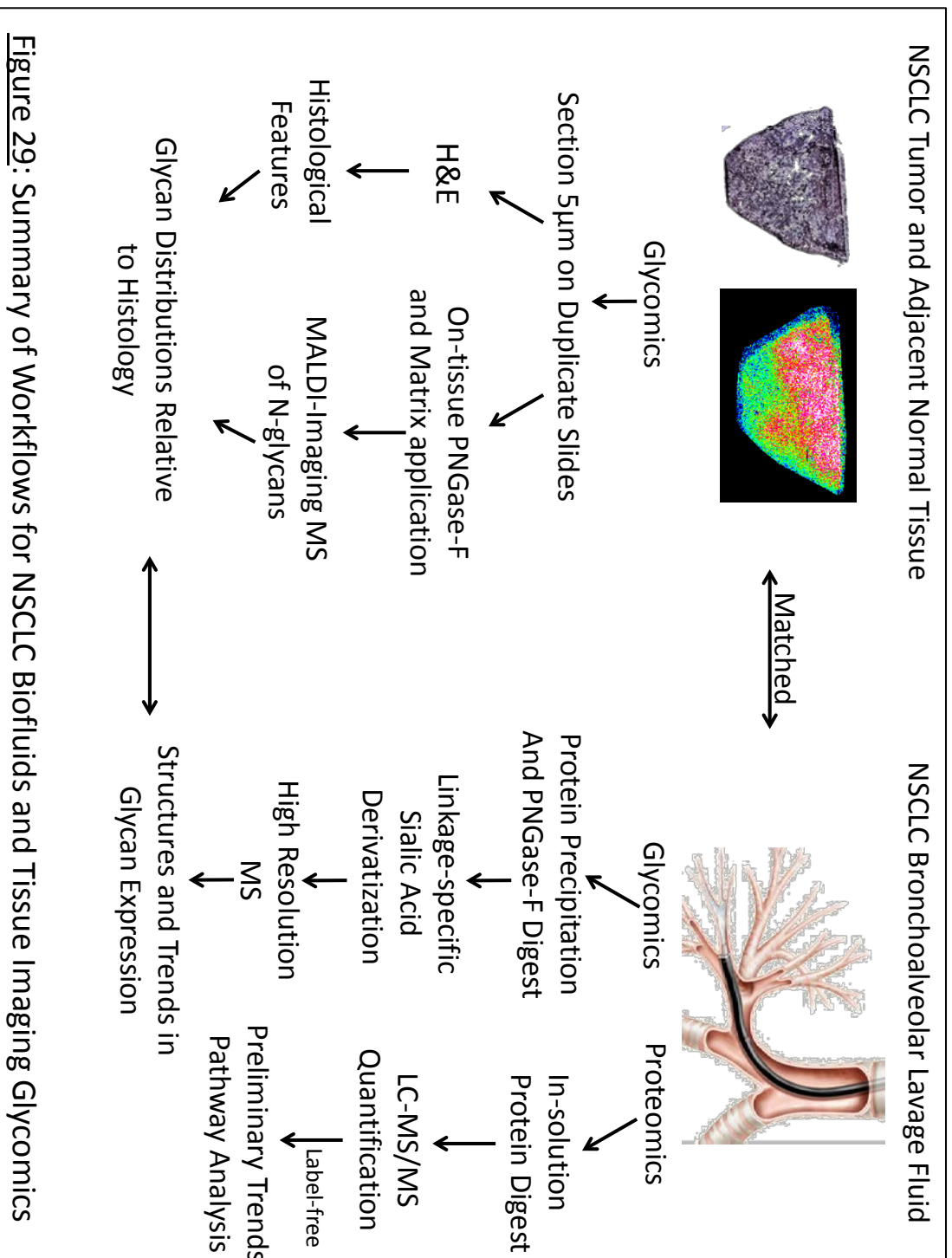
In relation to the MALDI tissue imaging approach presented, one of the limiting factors was the inability to attain spatial resolution at cellular and sub-cellular levels. This constraint has the potential to decrease the specificity of detected glycans because of the heterogeneous cell populations present in and adjacent to tumor tissues. Efforts to address this have successfully used an overlapping laser shot approach where repeated ablation of tissue resulted in improved resolution down to approximately 25 $\mu$ m(253). Alternatively, the use of smaller diameter lasers combined with a “transition geometry” method, where the MALDI laser is fired from behind the tissue, has also achieved higher definition in uncovering the distribution of tissue analytes at sub-micron levels(254). The feasibility of

incorporating these strategies into our workflow was principally constrained by the time and analytical requirements necessary to carry out these approaches. The selection of a raster width of 50 $\mu$ m was therefore an appropriate compromise to attain moderately high resolution images in a time frame conducive to high-throughput studies.

Another factor to consider for the glycan imaging studies was the discovery of inflammation and immune cell infiltrate within the tissues during histological assessment. This finding was unsurprising granted the established role for immune activation and inflammation in the pathogenesis of NSCLC(255). Importantly, several studies have taken advantage of the presence of such infiltrating cells to research the prognostic significance specific populations may confer (7,256). Further utility of our approach might therefore be achieved by incorporating a phenotypic evaluation of these infiltrating cells as an additional metric for prognostic considerations. Additionally, the discovery of immune cell and/or inflammatory-associated glycan epitopes from tissue imaging studies would be useful for this purpose, particularly if they could be translated into biofluids with relevance for clinical assays.

In summary, the assessment of a combined approach to identify N-glycans derived from matched lung tissues and proximal fluids was presented (**Figure 29**). Oligosaccharides were released from BAL and saliva pools and underwent a sialic acid linkage-specific derivatization strategy. This procedure serves to both protect from the loss of labile sialic acid residues during ionization and allow for discernment of linkage conformations. At the same time, tumor and normal lung tissues matched to the proximal fluids were sectioned for histological analysis and on-tissue acquisition of N-glycans. This approach facilitated discernment of the spatial distribution of N-glycan species to

relevant histological features of the tissue. All MS analyses were carried out on a MALDI-FT-ICR instrument, providing high-resolving power to differentiate a variety of oligomannose and complex glycans from each sample type. The implementation of these analytical strategies carried out on a set of matched samples was conducive to directly cross-referencing distinct glycan structures between tissues and biofluids. Therefore, ongoing studies focused to increasing the throughput, applying statistical analyses, and assessing reproducibility from these methods are likely to result in a whole new paradigm for approaches to glycomic biomarker discovery.



**Figure 29:** Summary of Workflows for NSCLC Biofluids and Tissue Imaging Glycomics

## Chapter 5

### Conclusions and Future Studies



## 5.1 Glycoproteomic and Glycomic Approaches to Evaluate TGF $\beta$ 1-induced effects on Sialoglycoproteins in a Cell Culture Model of NSCLC

### 5.1.1 Conclusions

We hypothesized that the development of mass spectrometry-based glycoproteomic and glycomic workflows could be used to identify detailed characteristics in sialylated glycoproteins in NSCLC. To test this, investigations were carried out in the H358 cells using an azido-sugar metabolic labeling, alkyne agarose bead enrichment, and LC-MS/MS quantification strategy. Differential capture of glycoproteins incorporating ManNAcAz following TGF $\beta$ 1 induction was assessed by both label-free and SILAC relative quantification. The double metabolic labeling strategy was also utilized to examine relative protein expression levels influenced by TGF $\beta$ 1. In addition, the utility of a secondary digest using PNGase-F was gauged for surveying N-linked glycosylation sites. Finally, the global N-glycome of TGF $\beta$ 1-induced H358 cells was acquired using an innovative pre-analytical strategy and MALDI-FT-ICR analysis. The application of these approaches revealed that:

1. We developed a robust method for the interrogation of sialylated glycoproteins in a cell culture model of NSCLC by mass spectrometry. A multi-faceted approach, including quantification of glycoproteins incorporating a sialic acid analog following the induction of TGF $\beta$ 1, elucidation of N-linked glycosylation sites, and structural evaluations of global N-linked glycans, was presented using the cell lysate fraction of H358s. Initial

assessments of these strategies also demonstrated feasibility for the methods with the GalNAcAz label and cell culture supernatant fractions.

2. H358 cells efficiently incorporate the ManNAcAz label onto nascent glycoproteins as demonstrated by WB analysis of cell lysates using biotin-alkyne/streptavidin-IR dye. The induction of TGF $\beta$ 1 appears to increase the presentation of the analog dramatically.

3. ManNAcAz labeling and Click-chemistry alkyne-agarose bead enrichment are an efficient means to selectively study low-abundance glycoproteins. This was evidenced by the high proportion of glycoproteins identified in SILAC+Click experiments (76%) relative to SILAC-Only experiments (11%).

4. The induction of TGF $\beta$ 1 expression in ManNAcAz-labeled cells resulted in significant differences in the captured quantities of glycoproteins relevant to the acquisition of invasive phenotypes. Supporting this was KEGG enrichment analysis of proteins captured at significantly different levels representing cell adhesion, extracellular matrix, and cytoskeletal-related functions.

5. The inclusion of SILAC with the azido-sugar metabolic labeling procedures provided a reliable means for quantifying the large number of glycoproteins identified. Even after the application of stringent post-processing parameters and replicate trimming, over 390 proteins were quantified from Click enriched samples and 700 from non-enriched.

6. The double metabolic labeling strategy provided limited information regarding how TGF $\beta$ 1-influenced protein expression levels were contributing to differential capture quantities. This was evident from the low number of proteins shared between the SILAC-Only and SILAC+Click fractions (91 total). Furthermore, only 5 proteins were found to be significantly enriched in the SILAC+Click while not significant in SILAC-Only at

$BH_p = 0.05$  and  $0.10$ . A more prudent approach to assess protein expression may be to perform WB analysis on selected glycoproteins relevant to invasive biological processes.

7. The utility of the alkyne-agarose bead enrichment procedure can be extended by the application of a secondary digest to release bound N-linked glycopeptides. This strategy is useful for verification of N-linked sites where ManNAcAz was incorporated and may reveal additional glycoproteins not detected by the initial digest.

8. MALDI-FT-ICR analysis of H358 N-linked glycoforms, using two derivatization methods, identified an array of the most prominent oligosaccharide structures from TGF $\beta$  and EV-induced cells. The permethylation approach was useful for identification of oligomannose and tri- and tetra-antennary complex glycans. The linkage-specific esterification method resulted in elucidation of anomeric configurations of sialic acid on several bi-antennary complex glycans.

### 5.1.2 Limitations of Azido-sugar Metabolic Labeling Studies

The azido-sugar metabolic labeling, alkyne-agarose bead enrichment and LC-MS/MS approach is a powerful tool for analyzing characteristics of sialylated glycoproteins, however there were some inherent technical and conceptual limitations associated with this strategy. These shortcomings are introduced here and discussed in more detail in the future studies section (5.1.3).

One limitation is the inability to release bound glycans from the alkyne agarose beads for structural evaluations. This limitation is due to the covalent nature of the Click reaction between the azide and alkyne functional groups. The strength of this reaction is an advantage for proteomic evaluations of captured glycoproteins, because it allows stringent washes of the beads to remove non-specifically bound molecules. At the same

time, however, the glycans which incorporated the azido-sugar remain permanently bound to the beads and there is currently no method available to release them. Therefore, the presented approach provides proteomic and N-glycan site data on captured glycoproteins without concomitant information on the glycan structures which incorporated the analog sugars.

Another drawback is the lack of information that can be obtained regarding O-linked glycosylation sites. This weakness is more related to the lack of an enzyme for O-glycan release from the alkyne beads than it is to the azido-sugar labeling and enrichment methodology. Further adding to this constraint is the lack of an amino acid consensus sequence for the attachment of O-linked sugars. Peptides attached to the beads linked via an O-glycosylation linkage could be released by beta-elimination, but there is no way to verify what site was modified, or how many sites were modified, for a given protein. Therefore, the presented approach is currently only amenable to identifying sites of N-linked glycosylation where the azido sugars are being incorporated, without equivalent information on sites of O-linked glycosylation.

Another weakness related to N-linked glycosylation site identification is the inability to differentiate random deamidations on asparagine residues from PNGase-F-mediated deamidations. Random deamidations may occur on asparagine or glutamine due to proteolytic degradation or because of side-chain reactions with glycine or serine/threonine residues. As a result, it is possible that some of the N-linked glycan sites identified were result chance occurrences; however the high proportion of peptides in which Asn was deamidated within an N-linked consensus sequence argues more for a PNGase-F specific deamidation.

Conceptual constraints also limit the relevance of information obtained in the presented approach. For example, in our demonstration of the methods, only one cell model of lung cancer was assessed. In addition, a single time-point of gene-induction and only one mechanism for inducing invasive phenotypes was carried out during experimental procedures. Therefore the findings presented do not adequately represent all relevant changes in sialylation that occur during the entire time-course of TGF $\beta$ 1-induced NSCLC progression; rather, our method provided a “snapshot” of sialylation changes occurring in a limited representation of the disease.

Another conceptual concern relates to the inability of the presented strategy to measure absolute protein and sialylation changes between experimental groups. Because the presented approach is based on relative quantification of captured glycoproteins, it is possible that differences in sialylation were missed when a given glycoprotein’s incorporation of the azido sugar was in opposition to changes in protein quantity. For example, if TGF $\beta$ 1 caused increased sialylation while simultaneously decreasing protein quantity for a given glycoprotein (relative to empty vector cells), then no difference in captured quantity would be seen between the groups. Therefore, the presented method is limited in that some of the differences in sialylation were likely over- or under-estimated.

### 5.1.3 Future Studies

The goal of this work was to develop methods and evaluate experimental approaches tailored to future glycoproteomic/glycomic biomarker studies in NSCLC. The initial experiments using azido sugar labeling and Click chemistry enrichment with label-free quantification demonstrated feasibility for the use of both ManNAcAz and GalNAcAz labels in cell lysates and supernatant fractions. However statistical analysis

and subsequent investigations were carried out only for ManNAcAz-labeled cell lysate fractions following TGF $\beta$ 1 induction. Therefore a more in-depth investigation into cell culture supernatant glycoproteins (i.e. secreted products) may be beneficial for translating in vitro findings into frequently used clinical fluids. An added benefit for this type of approach is that this method has previously been shown to maintain selectivity for low abundance glycoproteins without the complete removal of serum supplements from cell culture media(194). Furthermore, we have carried out preliminary assessments of this approach for the detection of exosomal glycoproteins (data not shown). These non-classical secreted vesicles have shown promise as a cancer biomarker discovery source as they carry tumor-specific protein products and may be isolated from multiple biological fluids(257,258). Therefore we additionally propose to implement the SILAC double metabolic labeling strategy for the quantification of exosomal glycoprotein cargo as a future direction.

Other variables to the presented strategy that will be addressed by future studies include a dedicated approach centered on the use of other sugar analogs, such as those for Fucose, GalNAc, or GlcNAc. These studies may provide a more comprehensive view of how other classes of glycoproteins respond to specific stimuli such as TGF $\beta$ 1. As well, for increasing the likelihood that in vitro findings will be representative of in vivo disease, additional studies would benefit from applying these methods to multiple cell models. For the presented work, parallel investigations using A549 or HCC4006 cells may have highlight inconsistencies or uncover common trends with the H358 model. Similar to this, utilization of alternative mechanisms for the production of specific phenotypic properties will be critical if claims will be made that observed glycosylation

changes are truly of a consequence these processes. For example, as an alternative to the use of TGF $\beta$ 1 to induce invasive characteristics, future studies might use hepatocyte growth factor or epidermal growth factor treatments, knockdowns of miR200, or the modification of E-cadherin to achieve this outcome.

The primary workflow utilized for azido sugar labeling and TGF $\beta$ 1 induction (**Figure 2**) was based on a previously optimized procedure described by Yang et al with slight modifications (193). Specifically, cells are incubated with the analog sugars for 48 hours, followed by a media change and the addition of doxycycline for 24 hours prior to collection of lysates and cell culture media. In order to gain a better perspective as to how sialylation is affected by TGF $\beta$ 1 expression, different gene induction time-points should be utilized. For example, studies with extended exposure to TGF $\beta$ 1 or experimental approaches where analog sugar metabolic labeling occurs after doxycycline treatment would be worthwhile endeavors.

In chapter 3, we utilized a broad-approach to identify large populations of glycoproteins potentially undergoing alternative sialylation in response to TGF $\beta$ 1 stimulation. The logical next step from these investigations would be to critically evaluate the individual proteins identified from this work. Future studies will therefore employ approaches such as immunoprecipitation to selectively focus on the glycosylation of specific targets. Glycoproteins captured at significantly different levels from SILAC+Click will be individually evaluated for protein expression levels by WB prior to moving forward. Mass spectrometry analyses such as HCD-PD-ETD will be employed to simultaneously gauge individual glycans and their site localizations on a given glycoprotein of interest. Preliminary studies in our lab, in collaboration with Dr. Robert

Gemmill, have already begun assessing the utility for this type of approach on putative targets such as Neuropilin-2. Furthermore, PNGase-F based release of N-glycans and utilization of linkage-specific derivatization methodology can be used as a parallel approach to both verify oligosaccharide structures and attain sialic acid linkage information from the individual targets.

Modifications to our strategy for glycosylation site assessment would also be an appropriate endeavor for future work. One of the central limitations to the method presented was that only N-linked glycopeptides were released. Implementing an approach for release of O-linked glycopeptides, such as beta elimination or hydrazinolysis, would be useful for assessing additional glycoproteins incorporating the analog. For O-linked site elucidation, difficulties arise because of the lack of consensus sequence for these glycans. Therefore, for this specific goal, exoglycosidase release of bead-bound glycopeptides followed by intact glycopeptide analysis would be the best approach.

Another avenue worth exploring would be accounting for the SILAC labels incorporated into peptides released from the alkyne beads by PNGase-F. This strategy would allow us to record the ratios of peptides derived from either experimental condition (TGF $\beta$  or EV) because the samples are mixed at 1:1 ratio prior to alkyne bead enrichment. This strategy has the potential to provide quantitative information on sialylation occurring at distinct glycosylation sites.

Quantitative glycomic studies may also be carried out using the cell culture model of NSCLC. These investigations have the potential to uncover trends in the expression of specific glycan structures which warrant further exploration in clinically-relevant samples (for example, by using lectin affinity enrichment to pulldown sialylated glycans of



interest in human biofluids/tissues). In an experimental approach similar to SILAC, isotopic detection of amino sugars with glutamine (iDAWG), has shown promise for the quantification of N- and O-glycan structures (259,260). This method is based on the metabolic incorporation of isotopically labeled nitrogen into GlcNAc, GalNAc, and sialic acid monosaccharides, which all rely on the side chain of glutamine as their source of nitrogen for biosynthesis.

## 5.2 Glycomic Assessment of Human NSCLC Biofluids and Tissue

### 5.2.1 Conclusions

In aim 2, we tested the application of new glycomic mass spectrometry acquisition strategies in human biofluids and lung tissues to investigate their utility in clinically-relevant sample types. A set of matched proximal fluids and tissue representing cases of benign disease, adenocarcinoma, and squamous cell carcinoma were utilized for “proof-of-concept” experiments. The proteome of BAL fluids was characterized by LC-MS/MS and preliminary trends were highlighted between the diagnostic groups. N-linked glycans from pooled samples of BAL and saliva were subjected to a linkage-specific sialic acid esterification method prior to acquisition on a high resolution/high mass accuracy instrument. Normal and tumor lung tissue were sectioned for histological analysis and tissue imaging mass spectrometry of native N-glycans. The application of these methods demonstrated that:

1. A detailed characterization of N-linked glycan structures present in proximal fluids and tissues of lung cancer can be accomplished by using a combination of recently-developed MS glycomic strategies. These included methods for discerning the anomeric linkages of

sialylated N-glycans in biofluids and the application of tissue imaging mass spectrometry to correlate specific oligosaccharides to histological features of the tissues. The use of MALDI-FT-ICR for all MS acquisitions provided a high level of precision and mass accuracy for resolving structures in each starting material. The use of matched samples allowed for direct comparison of structures between the fluids/tissues, representing a whole new paradigm in biomarker discovery efforts. Furthermore, this was the first demonstration of feasibility for the use of BAL fluid as glycomic biomarker reservoir that we know of.

2. Bronchoalveolar lavage fluid contains a diverse set of proteins representing biological processes such as wound healing, inflammatory response and coagulation. Over 300 proteins were identified, approximately 35% of which matched the Uniprot keyword for glycoprotein. The application of label-free quantification was a convenient means for assessing preliminary trends in the quantity of proteins between cancer and benign samples.

3. BAL and saliva contain a repertoire of N-linked glycan structures composed of oligomannose and bi-/ tri-antennary complex glycans. The presented approach resulted in the identification of 24 glycans in BAL from  $m/z = 1200-3185$  and 16 glycans from saliva from  $m/z = 1400-2375$ . The derivatization method resolved the anomeric configurations for 11 and 5 sialylated glycans in BAL and saliva, respectively. A total of 11 glycans shared identification in both BAL and saliva in the  $m/z = 1400-2375$  range.

4. The application of imaging mass spectrometry methodology was amenable to defining spatial distribution of N-glycans within tumor and normal adjacent lung tissue. The compatibility of this strategy with concurrent histological assessment facilitated

comparison of oligosaccharide localization to pertinent histological features as well as cross-reference in BAL and saliva.

### 5.2.2 Limitations of Human Biofluids and Tissue Imaging Glycomic Studies

The implementation of sialic acid linkage specific esterification and MALDI-FT-ICR analysis of BAL and saliva was an efficient method for identifying sialylated glycan structures in these fluids. In addition, our workflow for MALDI imaging mass spectrometry of human lung facilitated cross-reference of prevalent N-glycans to defined histological features of the tissue, as well as against matched proximal fluids. While these methods and biofluids carry great potential for future biomarker studies, there are some limitations associated with the presented workflow. These potential drawbacks are introduced here and alternative approaches are discussed in the future studies section (5.2.3).

The first constraint is related to the depth of coverage for BAL proteomic studies. This data was acquired following an in-solution trypsin digest, linear-ion trap mass spectrometry, and label-free quantification of detected proteins. This was an appropriate strategy to initially catalogue the protein-carriers of glycan structures present in this fluid. However, high-abundance metabolic proteins in the fluid, and the relatively-low resolution MS instrument used for acquisitions, likely hindered the identification of lower-abundance species. Furthermore, the application of a label-free quantification strategy is not the preferred method for analyzing starting materials containing a wide range of protein concentrations.

The next limitation relates to the incompatibility of the sialic acid linkage specific esterification procedure with the tissue imaging workflow. When this procedure is carried

out on fluid-based starting materials, the final step before MS is to enrich for sialylated glycans using cotton-tip HILIC. However, because the workflow for imaging MS relies on the analysis of intact (non-homogenized) tissues, no equivalent enrichment can be carried out on these types of samples. Therefore the presented strategy for sialic acid linkage specific esterification cannot currently be applied to intact tissue samples prior to MALDI-imaging mass spectrometry, thus limiting the number of sialylated glycans detected in imaging analyses.

Another factor to consider is that the glycomic studies carried out on tissues/fluids were limited to acquisition of N-glycans only. Similar to the limitation stated in the azido-sugar labeling and enrichment experiments, this relates primarily to the lack of an enzyme which specifically releases O-linked glycans for structural elucidation of these oligosaccharides.

Conceptual limitations also hinder the ability to cross-reference findings between either of the presented approaches (in vitro and human proximal fluid/tissue studies). The azido-sugar labeling, enrichment, and LC-MS/MS method provides proteomic and glycosylation site information on sialoglycoproteins following 24 hours of TGF $\beta$ 1 stimulation. On the other hand, data obtained from the human tissues/proximal fluids methods provides information on the structure of sialylated N-glycans and their distribution in NSCLC tissues. Therefore each of the presented methods provides a distinct set of information which cannot be directly inter-related between sample types without the need for additional experiments.

### 5.2.3 Future Studies

The goal of the work presented in chapter 4 was to assess the implementation of innovative MS glycomic analyses on underutilized starting materials which may have merit for biomarker discovery efforts. Based on our success in identifying a variety of N-linked glycans in each of these substrates, the most obvious next step will be to increase the number of samples studied so that trends in expression of specific glycan structures can be statistically evaluated. These efforts are already underway and relevant discoveries will be correlated to outcome information also available from the lung cancer biospecimen resource network. In addition, these methods will be applied to a wider array of lung proximal fluids as they become available. For example, exhaled breath condensate and pleural fluid have both demonstrated promising results for proteomic studies of lung cancer(261,262).

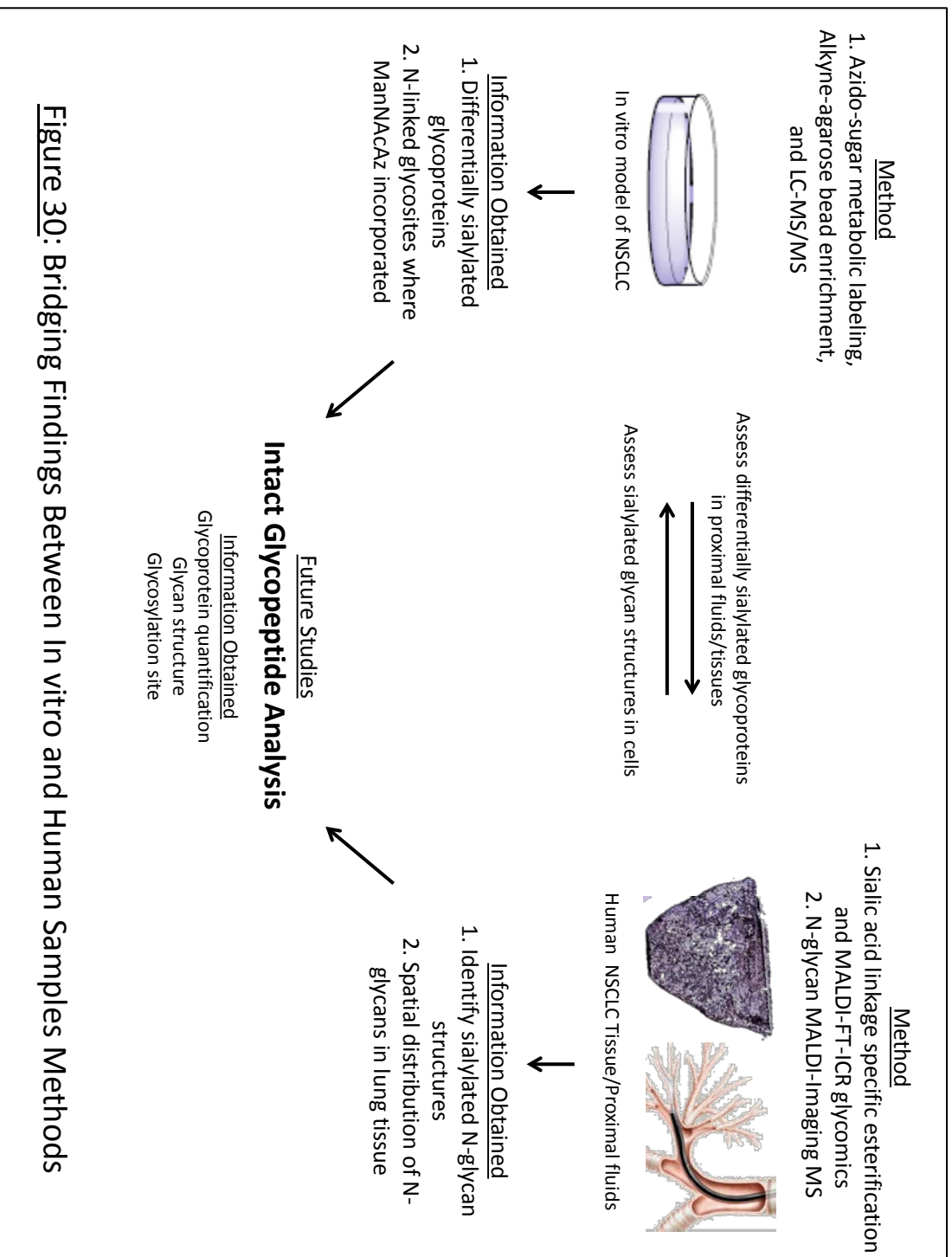
Another future endeavor will be to utilize an alternative means to quantify proteins present in the BAL fluid, such iTRAQ or TMT tags. Our study relied on LC-MS/MS and label-free quantification to catalogue the BAL proteome and explore preliminary trends in the expression of proteins between diagnostic groups. This approach was both rapid and suitable for this performing these experiments using limited sample volume. However as procurement of these samples increases, future work should be carried out using the most reliable quantitative methods for shotgun proteomic-type approaches.

For imaging studies of N-glycans, a reasonable goal for future studies will be to develop a method for on-tissue sialic acid linkage specific derivatization. This achievement would likely increase the number of sialylated glycans detected on-tissue

and thereby facilitate greater overlap of structures identified between matched samples. The primary difficulty in achieving this end thus far has been the preservation of spatial distribution of liberated glycans. As an alternative, tissue-imaging performed in negative ionization mode also has the potential to accomplish this feat. In addition to these goals, our workflow will also be modified to study glycan distribution following release with glycosidases such as sialidase or fucosidase. This strategy would be worthwhile to gain a better perspective of glycan diversity as it relates to tissue architecture. Furthermore, to address the spatial resolution limitation of imaging mass spectrometry, methods such as laser capture microdissection (LCM) could be implemented prior to MS analysis(263). For example, using LCM to selectively isolate transformed cells, without inclusion of adjacent heterogeneous cell populations, could serve to verify the specificity of cancer-associated glycan species. This approach has previously demonstrated utility in proteomic and glycomic investigations and therefore would be an appropriate follow-up to tissue imaging studies.

Finally, to address the conceptual limitation of not being able to directly compare findings between the in vitro and human biofluids/tissues methods, we have begun optimizing a strategy for intact glycopeptide analysis of biofluids and cell lysate (**Figure 30**). This method is based on a 2,2,2-trifluoroethanol acid and ammonium bicarbonate solubilization of membrane proteins, HILIC enrichment of glycopeptides, and HCD-PD-ETD acquisition on the OrbiTrap Elite instrument. Data from this technique will serve to verify glycan structures identified from MALDI-FT-ICR analyses as well as discern the specific glycosylation sites and protein carriers from which they were derived. By using an up-front immunoprecipitation or lectin affinity enrichment strategy prior to intact

glycopeptide analysis, glycoproteins and glycan structures of interest can be cross-referenced between sample types.



**Figure 30: Bridging Findings Between In vitro and Human Samples Methods**



## References

1. American Cancer Society. Lung Cancer (Non-Small Cell) [Internet]. 2014. Available from: <http://www.cancer.org/acs/groups/cid/documents/webcontent/003115-pdf.pdf>
2. American Cancer Society. Cancer Facts & Figures 2014 [Internet]. Atlanta: American Cancer Society; 2014. Available from: <http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf>
3. Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, et al. Non-small cell lung cancer. *J Natl Compr Cancer Netw JNCCN*. 2012 Oct 1;10(10):1236–71.
4. Horn L, Eisenberg R, Guis D, Kimmelschue K, Massion P, Putnam J, et al. Cancer of the Lung: Non-Small Cell Lung Cancer and Small Cell Lung Cancer. *Abeloff's Clinical Oncology*. 5th ed. Philadelphia: Elsevier Saunders; 2014. p. 1143–92.
5. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008 Jul 24;454(7203):436–44.
6. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002 Dec 19;420(6917):860–7.
7. O'Callaghan DS, O'Donnell D, O'Connell F, O'Byrne KJ. The role of inflammation in the pathogenesis of non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2010 Dec;5(12):2024–36.
8. Cooper WA, Lam DCL, O'Toole SA, Minna JD. Molecular biology of lung cancer. *J Thorac Dis*. 2013 Oct;5 Suppl 5:S479–90.
9. Finigan JH, Kern JA. Lung cancer screening: past, present and future. *Clin Chest Med*. 2013 Sep;34(3):365–71.
10. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011 Aug 4;365(5):395–409.
11. National Lung Screening Trial Research Team, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011 Jan;258(1):243–53.
12. Kerr KM, Bubendorf L, Edelman MJ, Marchetti A, Mok T, Novello S, et al. Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer. *Ann Oncol*. 2014 Sep 1;25(9):1681–90.

13. Mehan MR, Williams SA, Siegfried JM, Bigbee WL, Weissfeld JL, Wilson DO, et al. Validation of a blood protein signature for non-small cell lung cancer. *Clin Proteomics*. 2014;11(1):32.
14. Li QK, Gabrielson E, Askin F, Chan DW, Zhang H. Glycoproteomics using fluid-based specimens in the discovery of lung cancer protein biomarkers: Promise and challenge. *PROTEOMICS - Clin Appl*. 2013 Jan;7(1-2):55–69.
15. Li QK, Shah P, Li Y, Aiyetan PO, Chen J, Yung R, et al. Glycoproteomic analysis of bronchoalveolar lavage (BAL) fluid identifies tumor-associated glycoproteins from lung adenocarcinoma. *J Proteome Res*. 2013 Aug 2;12(8):3689–96.
16. Travis WD, Brambilla E, Riely GJ. New Pathologic Classification of Lung Cancer: Relevance for Clinical Practice and Clinical Trials. *J Clin Oncol*. 2013 Mar 10;31(8):992–1001.
17. Davidson MR, Gazdar AF, Clarke BE. The pivotal role of pathology in the management of lung cancer. *J Thorac Dis*. 2013 Oct;5 Suppl 5:S463–78.
18. Kumar V, Abbas A, Aster J, editors. *Lungs. Robbins basic pathology*. 9th ed. Philadelphia, PA: Elsevier/Saunders; 2013. p. 459–515.
19. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014 Jul 24;14(8):535–46.
20. Drilon A, Rekhtman N, Ladanyi M, Paik P. Squamous-cell carcinomas of the lung: emerging biology, controversies, and the promise of targeted therapy. *Lancet Oncol*. 2012 Oct;13(10):e418–26.
21. Li MO, Wan YY, Sanjabi S, Robertson A-KL, Flavell RA. Transforming Growth Factor- $\beta$  Regulation of Immune Responses. *Annu Rev Immunol*. 2006 Apr;24(1):99–146.
22. Massagué J. TGF $\beta$  signalling in context. *Nat Rev Mol Cell Biol*. 2012 Oct;13(10):616–30.
23. Bierie B, Moses HL. Tumour microenvironment: TGF $\beta$ : the molecular Jekyll and Hyde of cancer. *Nat Rev Cancer*. 2006 Jul;6(7):506–20.
24. Bierie B, Moses HL. Transforming growth factor beta (TGF- $\beta$ ) and inflammation in cancer. *Cytokine Growth Factor Rev*. 2010 Feb;21(1):49–59.
25. Ikushima H, Miyazono K. TGF $\beta$  signalling: a complex web in cancer progression. *Nat Rev Cancer*. 2010 Jun;10(6):415–24.
26. Daly AC, Randall RA, Hill CS. Transforming Growth Factor -Induced Smad1/5 Phosphorylation in Epithelial Cells Is Mediated by Novel Receptor Complexes and Is

- Essential for Anchorage-Independent Growth. *Mol Cell Biol.* 2008 Nov 15;28(22):6889–902.
27. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature.* 2003 Oct 9;425(6958):577–84.
  28. Heldin C-H, Vanlandewijck M, Moustakas A. Regulation of EMT by TGF $\beta$  in cancer. *FEBS Lett.* 2012 Jul;586(14):1959–70.
  29. Wang X-J, Han G, Owens P, Siddiqui Y, Li AG. Role of TGF $\beta$ -Mediated Inflammation in Cutaneous Wound Healing. *J Invest Dermatol Symp Proc.* 2006 Sep;11(1):112–7.
  30. Xie L, Law BK, Aakre ME, Edgerton M, Shyr Y, Bhowmick NA, et al. Transforming growth factor beta-regulated gene expression in a mouse mammary gland epithelial cell line. *Breast Cancer Res BCR.* 2003;5(6):R187–98.
  31. Zhang YE. Non-Smad pathways in TGF- $\beta$  signaling. *Cell Res.* 2009 Jan;19(1):128–39.
  32. Levy L, Hill C. Alterations in components of the TGF- $\beta$  superfamily signaling pathways in human cancer. *Cytokine Growth Factor Rev.* 2006 Feb;17(1-2):41–58.
  33. Tsushima H, Kawata S, Tamura S, Ito N, Shirai Y, Kiso S, et al. High levels of transforming growth factor beta 1 in patients with colorectal cancer: association with disease progression. *Gastroenterology.* 1996 Feb;110(2):375–82.
  34. Wikström P, Stattin P, Franck-Lissbrant I, Damber JE, Bergh A. Transforming growth factor beta1 is associated with angiogenesis, metastasis, and poor clinical outcome in prostate cancer. *The Prostate.* 1998 Sep 15;37(1):19–29.
  35. Walker RA, Dearing SJ. Transforming growth factor beta 1 in ductal carcinoma in situ and invasive carcinomas of the breast. *Eur J Cancer Oxf Engl* 1990. 1992;28(2-3):641–4.
  36. Friedman E, Gold LI, Klimstra D, Zeng ZS, Winawer S, Cohen A. High levels of transforming growth factor beta 1 correlate with disease progression in human colon cancer. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol.* 1995 Aug;4(5):549–54.
  37. Reiss M. TGF- $\beta$  and cancer. *Microbes Infect.* 1999 Dec;1(15):1327–47.
  38. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest.* 2009 Jun;119(6):1420–8.
  39. Willis BC, Borok Z. TGF-beta-induced EMT: mechanisms and implications for fibrotic lung disease. *AJP Lung Cell Mol Physiol.* 2007 Jun 29;293(3):L525–34.

40. Branton MH, Kopp JB. TGF-beta and fibrosis. *Microbes Infect Inst Pasteur*. 1999 Dec;1(15):1349–65.
41. Alcaraz A, Mrowiec A, Insausti CL, García-Vizcaíno EM, Ruiz-Canada C, López-Martínez MC, et al. Autocrine TGF- $\beta$  Induces Epithelial to Mesenchymal Transition in Human Amniotic Epithelial Cells. *Cell Transplant*. 2013 Aug 9;22(8):1351–67.
42. Ewen ME, Oliver CJ, Sluss HK, Miller SJ, Peeper DS. p53-dependent repression of CDK4 translation in TGF-beta-induced G1 cell-cycle arrest. *Genes Dev*. 1995 Jan 15;9(2):204–17.
43. Wahl SM, Allen JB, Weeks BS, Wong HL, Klotman PE. Transforming growth factor beta enhances integrin expression and type IV collagenase secretion in human monocytes. *Proc Natl Acad Sci U S A*. 1993 May 15;90(10):4577–81.
44. Freire-de-Lima L. Sweet and Sour: The Impact of Differential Glycosylation in Cancer Cells Undergoing Epithelial to Mesenchymal Transition. *Front Oncol* [Internet]. 2014 Mar 25 [cited 2014 Oct 29];4. Available from: <http://journal.frontiersin.org/Journal/10.3389/fonc.2014.00059/full>
45. Maupin KA, Sinha A, Eugster E, Miller J, Ross J, Paulino V, et al. Glycogene Expression Alterations Associated with Pancreatic Cancer Epithelial-Mesenchymal Transition in Complementary Model Systems. Chammas R, editor. *PLoS ONE*. 2010 Sep 27;5(9):e13002.
46. Zhang H, Meng F, Wu S, Kreike B, Sethi S, Chen W, et al. Engagement of I-branching  $\beta$ -1, 6-N-acetylglucosaminyltransferase 2 in breast cancer metastasis and TGF- $\beta$  signaling. *Cancer Res*. 2011 Jul 15;71(14):4846–56.
47. Pinho SS, Oliveira P, Cabral J, Carvalho S, Huntsman D, Gärtner F, et al. Loss and recovery of Mgat3 and GnT-III Mediated E-cadherin N-glycosylation is a mechanism involved in epithelial-mesenchymal-epithelial transitions. *PLoS One*. 2012;7(3):e33191.
48. Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. *Cell*. 2006 Sep 8;126(5):855–67.
49. Varki A, Freeze HH, Vacquier VD. Glycans in Development and Systemic Physiology. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2015 Jan 5]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1906/>
50. Tran DT, Ten Hagen KG. Mucin-type O-Glycosylation during Development. *J Biol Chem*. 2013 Mar 8;288(10):6921–9.

51. Dube DH, Bertozzi CR. Glycans in cancer and inflammation — potential for therapeutics and diagnostics. *Nat Rev Drug Discov.* 2005 Jun;4(6):477–88.
52. Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. *PROTEOMICS.* 2014 Mar;14(4-5):525–46.
53. Ono M, Hakomori S. Glycosylation defining cancer cell motility and invasiveness. *Glycoconj J.* 2004;20(1):71–8.
54. Taniguchi N, Miyoshi E, Ko JH, Ikeda Y, Ihara Y. Implication of N-acetylglucosaminyltransferases III and V in cancer: gene regulation and signaling mechanism. *Biochim Biophys Acta.* 1999 Oct 8;1455(2-3):287–300.
55. Xu Q, Isaji T, Lu Y, Gu W, Kondo M, Fukuda T, et al. Roles of N-Acetylglucosaminyltransferase III in Epithelial-to-Mesenchymal Transition Induced by Transforming Growth Factor 1 (TGF- 1) in Epithelial Cell Lines. *J Biol Chem.* 2012 May 11;287(20):16563–74.
56. Granovsky M, Fata J, Pawling J, Muller WJ, Khokha R, Dennis JW. Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nat Med.* 2000 Mar;6(3):306–12.
57. Freire-de-Lima L, Gelfenbeyn K, Ding Y, Mandel U, Clausen H, Handa K, et al. Involvement of O-glycosylation defining oncofetal fibronectin in epithelial-mesenchymal transition process. *Proc Natl Acad Sci.* 2011 Oct 25;108(43):17690–5.
58. Varki A, Lowe JB. Biological Roles of Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1897/>
59. Jamal BT, Nita-Lazar M, Gao Z, Amin B, Walker J, Kukuruzinska MA. N-glycosylation status of E-cadherin controls cytoskeletal dynamics through the organization of distinct  $\beta$ -catenin- and  $\gamma$ -catenin-containing AJs. *Cell Health Cytoskelet.* 2009 Sep 16;2009(1):67–80.
60. Freeze HH, Schachter H. Genetic Disorders of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2015 Jan 7]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1939/>
61. Jaeken J, Matthijs G. Congenital disorders of glycosylation: a rapidly expanding disease family. *Annu Rev Genomics Hum Genet.* 2007;8:261–78.

62. Haltiwanger RS, Lowe JB. Role of glycosylation in development. *Annu Rev Biochem.* 2004;73:491–537.
63. Varki A, Kannagi R, Toole BP. Glycosylation Changes in Cancer. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1963/>
64. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol.* 2012 Jun 22;13(7):448–62.
65. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1921/>
66. Varki A, Esko JD, Colley KJ. Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1926/>
67. Bertozzi CR, Rabuka D. Structural Basis of Glycan Diversity. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1955/>
68. Esko JD, Kimata K, Lindahl U. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2015 Jan 7]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1900/>
69. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1917/>
70. Freeze HH, Esko JD, Parodi AJ. Glycans in Glycoprotein Quality Control. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor

Laboratory Press; 2009 [cited 2014 Oct 30]. Available from:  
<http://www.ncbi.nlm.nih.gov/books/NBK1958/>

71. Brockhausen I, Schachter H, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from:  
<http://www.ncbi.nlm.nih.gov/books/NBK1896/>
72. Desai PR. Immunoreactive T and Tn antigens in malignancy: role in carcinoma diagnosis, prognosis, and immunotherapy. *Transfus Med Rev*. 2000 Oct;14(4):312–25.
73. Stanley P, Cummings RD. Structures Common to Different Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from:  
<http://www.ncbi.nlm.nih.gov/books/NBK1892/>
74. Gendler SJ. MUC1, the renaissance molecule. *J Mammary Gland Biol Neoplasia*. 2001 Jul;6(3):339–53.
75. Du J, Meledeo MA, Wang Z, Khanna HS, Paruchuri VDP, Yarema KJ. Metabolic glycoengineering: sialic acid and beyond. *Glycobiology*. 2009 Dec;19(12):1382–401.
76. Varki A, Schauer R. Sialic Acids. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology* [Internet]. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 [cited 2014 Oct 30]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1920/>
77. Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P. The human sialyltransferase family. *Biochimie*. 2001 Aug;83(8):727–37.
78. Schultz MJ, Swindall AF, Bellis SL. Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metastasis Rev*. 2012 Dec;31(3-4):501–18.
79. Mühlhoff M, Rollenhagen M, Werneburg S, Gerardy-Schahn R, Hildebrandt H. Polysialic acid: versatile modification of NCAM, SynCAM 1 and neuropilin-2. *Neurochem Res*. 2013 Jun;38(6):1134–43.
80. Büll C, Stoel MA, den Brok MH, Adema GJ. Sialic acids sweeten a tumor's life. *Cancer Res*. 2014 Jun 15;74(12):3199–204.
81. Dall'Olio F, Chiricolo M. Sialyltransferases in cancer. *Glycoconj J*. 2001 Dec;18(11-12):841–50.

82. Samraj AN, Läubli H, Varki N, Varki A. Involvement of a non-human sialic Acid in human cancer. *Front Oncol.* 2014;4:33.
83. Schreiber SC, Giehl K, Kastilan C, Hasel C, Mühlenhoff M, Adler G, et al. Polysialylated NCAM represses E-cadherin-mediated cell-cell adhesion in pancreatic tumor cells. *Gastroenterology.* 2008 May;134(5):1555–66.
84. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep.* 2006 Jun;7(6):599–604.
85. Guo H-B, Lee I, Kamar M, Akiyama SK, Pierce M. Aberrant N-glycosylation of beta1 integrin causes reduced alpha5beta1 integrin clustering and stimulates cell migration. *Cancer Res.* 2002 Dec 1;62(23):6837–45.
86. Renkonen J, Paavonen T, Renkonen R. Endothelial and epithelial expression of sialyl Lewis(x) and sialyl Lewis(a) in lesions of breast carcinoma. *Int J Cancer J Int Cancer.* 1997 Jun 20;74(3):296–300.
87. Miyake M, Taki T, Hitomi S, Hakomori S. Correlation of expression of H/Le(y)/Le(b) antigens with survival in patients with carcinoma of the lung. *N Engl J Med.* 1992 Jul 2;327(1):14–8.
88. Nakamori S, Kameyama M, Imaoka S, Furukawa H, Ishikawa O, Sasaki Y, et al. Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study. *Cancer Res.* 1993 Aug 1;53(15):3632–7.
89. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, et al. Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clin Chem.* 2010 Feb 1;56(2):223–36.
90. Miyoshi E, Nishikawa A, Ihara Y, Saito H, Uozumi N, Hayashi N, et al. Transforming growth factor beta up-regulates expression of the N-acetylglucosaminyltransferase V gene in mouse melanoma cells. *J Biol Chem.* 1995 Mar 17;270(11):6216–20.
91. Pinho SS, Seruca R, Gärtner F, Yamaguchi Y, Gu J, Taniguchi N, et al. Modulation of E-cadherin function and dysfunction by N-glycosylation. *Cell Mol Life Sci.* 2011 Mar;68(6):1011–20.
92. Partridge EA. Regulation of Cytokine Receptors by Golgi N-Glycan Processing and Endocytosis. *Science.* 2004 Oct 1;306(5693):120–4.
93. Zhao Y-Y, Takahashi M, Gu J-G, Miyoshi E, Matsumoto A, Kitazume S, et al. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci.* 2008 Jul;99(7):1304–10.



94. Abbott KL, Nairn AV, Hall EM, Horton MB, McDonald JF, Moremen KW, et al. Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. *PROTEOMICS*. 2008 Aug;8(16):3210–20.
95. Ishibashi K, Nishikawa A, Hayashi N, Kasahara A, Sato N, Fujii S, et al. N-acetylglucosaminyltransferase III in human serum, and liver and hepatoma tissues: increased activity in liver cirrhosis and hepatoma patients. *Clin Chim Acta Int J Clin Chem*. 1989 Dec 15;185(3):325–32.
96. Yoshimura M, Nishikawa A, Ihara Y, Taniguchi S, Taniguchi N. Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection. *Proc Natl Acad Sci U S A*. 1995 Sep 12;92(19):8754–8.
97. Yoshimura M, Ihara Y, Matsuzawa Y, Taniguchi N. Aberrant glycosylation of E-cadherin enhances cell-cell binding to suppress metastasis. *J Biol Chem*. 1996 Jun 7;271(23):13811–5.
98. Itzkowitz SH, Yuan M, Montgomery CK, Kjeldsen T, Takahashi HK, Bigbee WL, et al. Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res*. 1989 Jan 1;49(1):197–204.
99. Ju T, Aryal RP, Stowell CJ, Cummings RD. Regulation of protein O-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *J Cell Biol*. 2008 Aug 11;182(3):531–42.
100. Radhakrishnan P, Dabelsteen S, Madsen FB, Francavilla C, Kopp KL, Steentoft C, et al. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc Natl Acad Sci*. 2014 Sep 30;111(39):E4066–75.
101. Ju T, Lanneau GS, Gautam T, Wang Y, Xia B, Stowell SR, et al. Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res*. 2008 Mar 15;68(6):1636–46.
102. Marcos NT, Bennett EP, Gomes J, Magalhaes A, Gomes C, David L, et al. ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci Elite Ed*. 2011;3:1443–55.
103. Vázquez-Martín C, Cuevas E, Gil-Martín E, Fernández-Briera A. Correlation analysis between tumor-associated antigen sialyl-Tn expression and ST6GalNAc I activity in human colon adenocarcinoma. *Oncology*. 2004;67(2):159–65.
104. Yang JM, Byrd JC, Siddiki BB, Chung YS, Okuno M, Sowa M, et al. Alterations of O-glycan biosynthesis in human colon cancer tissues. *Glycobiology*. 1994 Dec;4(6):873–84.

105. Ogata S, Maimonis PJ, Itzkowitz SH. Mucins bearing the cancer-associated sialosyl-Tn antigen mediate inhibition of natural killer cell cytotoxicity. *Cancer Res.* 1992 Sep 1;52(17):4741–6.
106. Van de Wiel-van Kemenade E, Ligtenberg MJ, de Boer AJ, Buijs F, Vos HL, Melief CJ, et al. Episialin (MUC1) inhibits cytotoxic lymphocyte-target cell interaction. *J Immunol Baltim Md 1950.* 1993 Jul 15;151(2):767–76.
107. Van Vliet SJ, Paessens LC, Broks-van den Berg VCM, Geijtenbeek TBH, van Kooyk Y. The C-Type Lectin Macrophage Galactose-Type Lectin Impedes Migration of Immature APCs. *J Immunol.* 2008 Sep 1;181(5):3148–55.
108. Glinsky VV, Glinsky GV, Rittenhouse-Olson K, Huflejt ME, Glinskii OV, Deutscher SL, et al. The role of Thomsen-Friedenreich antigen in adhesion of human breast and prostate cancer cells to the endothelium. *Cancer Res.* 2001 Jun 15;61(12):4851–7.
109. Clement M. Expression of sialyl-Tn epitopes on beta1 integrin alters epithelial cell phenotype, proliferation and haptotaxis. *J Cell Sci.* 2004 Oct 1;117(21):5059–69.
110. Julien S. ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumorigenicity. *Glycobiology.* 2005 Jul 13;16(1):54–64.
111. Kannagi R. Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression-The Warburg effect revisited. *Glycoconj J.* 2004;20(5):353–64.
112. Läubli H, Borsig L. Selectins promote tumor metastasis. *Semin Cancer Biol.* 2010 Jun;20(3):169–77.
113. Koike T, Kimura N, Miyazaki K, Yabuta T, Kumamoto K, Takenoshita S, et al. Hypoxia induces adhesion molecules on cancer cells: A missing link between Warburg effect and induction of selectin-ligand carbohydrates. *Proc Natl Acad Sci U S A.* 2004 May 25;101(21):8132–7.
114. Brockhausen I, Yang JM, Burchell J, Whitehouse C, Taylor-Papadimitriou J. Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. *Eur J Biochem FEBS.* 1995 Oct 15;233(2):607–17.
115. Altschuler Y, Kinlough CL, Poland PA, Bruns JB, Apodaca G, Weisz OA, et al. Clathrin-mediated endocytosis of MUC1 is modulated by its glycosylation state. *Mol Biol Cell.* 2000 Mar;11(3):819–31.
116. Rahn JJ, Chow JW, Horne GJ, Mah BK, Emerman JT, Hoffman P, et al. MUC1 mediates transendothelial migration in vitro by ligating endothelial cell ICAM-1. *Clin Exp Metastasis.* 2005;22(6):475–83.

117. Winter JM, Tang LH, Klimstra DS, Brennan MF, Brody JR, Rocha FG, et al. A Novel Survival-Based Tissue Microarray of Pancreatic Cancer Validates MUC1 and Mesothelin as Biomarkers. Andre F, editor. PLoS ONE. 2012 Jul 6;7(7):e40157.
118. Duffy MJ, Evoy D, McDermott EW. CA 15-3: Uses and limitation as a biomarker for breast cancer. Clin Chim Acta. 2010 Dec;411(23-24):1869–74.
119. Smagghe BJ, Stewart AK, Carter MG, Shelton LM, Bernier KJ, Hartman EJ, et al. MUC1\* Ligand, NM23-H1, Is a Novel Growth Factor That Maintains Human Stem Cells in a More Naïve State. Tang YL, editor. PLoS ONE. 2013 Mar 7;8(3):e58601.
120. Seales EC, Jurado GA, Singhal A, Bellis SL. Ras oncogene directs expression of a differentially sialylated, functionally altered  $\beta$ 1 integrin. Oncogene. 2003 Oct 16;22(46):7137–45.
121. Uemura T, Shiozaki K, Yamaguchi K, Miyazaki S, Satomi S, Kato K, et al. Contribution of sialidase NEU1 to suppression of metastasis of human colon cancer cells through desialylation of integrin  $\beta$ 4. Oncogene. 2009 Mar 5;28(9):1218–29.
122. Seales EC. Hypersialylation of  $\beta$ 1 Integrins, Observed in Colon Adenocarcinoma, May Contribute to Cancer Progression by Up-regulating Cell Motility. Cancer Res. 2005 Jun 1;65(11):4645–52.
123. Bellis SL. Variant glycosylation: an underappreciated regulatory mechanism for beta1 integrins. Biochim Biophys Acta. 2004 May 27;1663(1-2):52–60.
124. Gu J, Taniguchi N. Regulation of integrin functions by N-glycans. Glycoconj J. 2004;21(1-2):9–15.
125. Pretzlaff RK, Xue VW, Rowin ME. Sialidase treatment exposes the beta1-integrin active ligand binding site on HL60 cells and increases binding to fibronectin. Cell Adhes Commun. 2000;7(6):491–500.
126. Semel AC, Seales EC, Singhal A, Eklund EA, Colley KJ, Bellis SL. Hyposialylation of integrins stimulates the activity of myeloid fibronectin receptors. J Biol Chem. 2002 Sep 6;277(36):32830–6.
127. Woodard-Grice AV, McBrayer AC, Wakefield JK, Zhuo Y, Bellis SL. Proteolytic shedding of ST6Gal-I by BACE1 regulates the glycosylation and function of alpha4beta1 integrins. J Biol Chem. 2008 Sep 26;283(39):26364–73.
128. Pocheć E, Lityńska A, Amoresano A, Casbarra A. Glycosylation profile of integrin alpha 3 beta 1 changes with melanoma progression. Biochim Biophys Acta. 2003 Dec 7;1643(1-3):113–23.

129. Tangvoranuntakul P, Gagneux P, Diaz S, Bardor M, Varki N, Varki A, et al. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci*. 2003 Oct 14;100(21):12045–50.
130. Hildebrandt H, Mühlenhoff M, Gerardy-Schahn R. Polysialylation of NCAM. *Adv Exp Med Biol*. 2010;663:95–109.
131. Bax M, van Vliet SJ, Litjens M, García-Vallejo JJ, van Kooyk Y. Interaction of Polysialic Acid with CCL21 Regulates the Migratory Capacity of Human Dendritic Cells. Neyrolles O, editor. *PLoS ONE*. 2009 Sep 14;4(9):e6987.
132. Hildebrandt H, Becker C, Glüer S, Rösner H, Gerardy-Schahn R, Rahmann H. Polysialic acid on the neural cell adhesion molecule correlates with expression of polysialyltransferases and promotes neuroblastoma cell growth. *Cancer Res*. 1998 Feb 15;58(4):779–84.
133. Roth J, Zuber C, Wagner P, Taatjes DJ, Weisgerber C, Heitz PU, et al. Reexpression of poly(sialic acid) units of the neural cell adhesion molecule in Wilms tumor. *Proc Natl Acad Sci U S A*. 1988 May;85(9):2999–3003.
134. Tanaka F, Otake Y, Nakagawa T, Kawano Y, Miyahara R, Li M, et al. Expression of polysialic acid and STX, a human polysialyltransferase, is correlated with tumor progression in non-small cell lung cancer. *Cancer Res*. 2000 Jun 1;60(11):3072–80.
135. Trouillas J, Daniel L, Guigard M-P, Tong S, Gouvernet J, Jouanneau E, et al. Polysialylated neural cell adhesion molecules expressed in human pituitary tumors and related to extrasellar invasion. *J Neurosurg*. 2003 May;98(5):1084–93.
136. Al-Saraireh YMJ, Sutherland M, Springett BR, Freiburger F, Ribeiro Morais G, Loadman PM, et al. Pharmacological inhibition of polysialyltransferase ST8Siall modulates tumour cell migration. *PLoS One*. 2013;8(8):e73366.
137. Packer NH, von der Lieth C-W, Aoki-Kinoshita KF, Lebrilla CB, Paulson JC, Raman R, et al. Frontiers in glycomics: Bioinformatics and biomarkers in disease An NIH White Paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006). *PROTEOMICS*. 2008 Jan;8(1):8–20.
138. Hayes DF, Zurawski VR, Kufe DW. Comparison of circulating CA15-3 and carcinoembryonic antigen levels in patients with breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 1986 Oct;4(10):1542–50.
139. Stieber P, Molina R, Chan DW, Fritsche HA, Beyrau R, Bonfrer JMG, et al. Clinical evaluation of the Elecsys CA 15-3 test in breast cancer patients. *Clin Lab*. 2003;49(1-2):15–24.

140. Stieber P, Molina R, Chan DW, Fritsche HA, Beyrau R, Bonfrer JM, et al. Evaluation of the analytical and clinical performance of the Elecsys CA 15-3 immunoassay. *Clin Chem*. 2001 Dec;47(12):2162–4.
141. Bon GG, von Mensdorff-Pouilly S, Kenemans P, van Kamp GJ, Verstraeten RA, Hilgers J, et al. Clinical and technical evaluation of ACS BR serum assay of MUC1 gene-derived glycoprotein in breast cancer, and comparison with CA 15-3 assays. *Clin Chem*. 1997 Apr;43(4):585–93.
142. Sturgeon CM, Duffy MJ, Stenman U-H, Lilja H, Brünner N, Chan DW, et al. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clin Chem*. 2008 Dec;54(12):e11–79.
143. Huang Z-B. Prognostic value of preoperative serum tumor markers in gastric cancer. *World J Clin Oncol*. 2014;5(2):170.
144. Ballehaninna UK, Chamberlain RS. Serum CA 19-9 as a Biomarker for Pancreatic Cancer—A Comprehensive Review. *Indian J Surg Oncol*. 2011 Jun;2(2):88–100.
145. Świdarska M, Choromańska B, Dąbrowska E, Konarzewska-Duchnowska E, Choromańska K, Szczurko G, et al. Review The diagnostics of colorectal cancer. *Współczesna Onkol*. 2014;1:1–6.
146. Koprowski H, Steplewski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P. Colorectal carcinoma antigens detected by hybridoma antibodies. *Somatic Cell Genet*. 1979 Nov;5(6):957–71.
147. Tessler DA, Catanzaro A, Velanovich V, Havstad S, Goel S. Predictors of cancer in patients with suspected pancreatic malignancy without a tissue diagnosis. *Am J Surg*. 2006 Feb;191(2):191–7.
148. Kim YC, Kim HJ, Park JH, Park DI, Cho YK, Sohn CI, et al. Can preoperative CA19-9 and CEA levels predict the resectability of patients with pancreatic adenocarcinoma? *J Gastroenterol Hepatol*. 2009 Dec;24(12):1869–75.
149. Safi F, Schlosser W, Kolb G, Beger HG. Diagnostic value of CA 19-9 in patients with pancreatic cancer and nonspecific gastrointestinal symptoms. *J Gastrointest Surg Off J Soc Surg Aliment Tract*. 1997 Apr;1(2):106–12.
150. Zhang S, Wang Y-M, Sun C-D, Lu Y, Wu L-Q. Clinical value of serum CA19-9 levels in evaluating resectability of pancreatic carcinoma. *World J Gastroenterol WJG*. 2008 Jun 21;14(23):3750–3.
151. Halloran CM, Ghaneh P, Connor S, Sutton R, Neoptolemos JP, Raraty MGT. Carbohydrate antigen 19.9 accurately selects patients for laparoscopic assessment

- to determine resectability of pancreatic malignancy. *Br J Surg*. 2008 Apr;95(4):453–9.
152. Waraya M, Yamashita K, Katagiri H, Ishii K, Takahashi Y, Furuta K, et al. Preoperative serum CA19-9 and dissected peripancreatic tissue margin as determiners of long-term survival in pancreatic cancer. *Ann Surg Oncol*. 2009 May;16(5):1231–40.
153. Kim J-E, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol*. 2004 Feb;19(2):182–6.
154. Bast RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest*. 1981 Nov;68(5):1331–7.
155. Munkarah A, Chatterjee M, Tainsky MA. Update on ovarian cancer screening. *Curr Opin Obstet Gynecol*. 2007 Feb;19(1):22–6.
156. Zhang Z, Bast RC, Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res*. 2004 Aug 15;64(16):5882–90.
157. Understanding CA 125 Levels: A Guide for Ovarian Cancer Patients [Internet]. Foundation for Women’s Cancer; 2011. Available from: <http://www.foundationforwomenscancer.org/wp-content/uploads/CA125levels.pdf>
158. Skates SJ, Mai P, Horick NK, Piedmonte M, Drescher CW, Isaacs C, et al. Large prospective study of ovarian cancer screening in high-risk women: CA125 cut-point defined by menopausal status. *Cancer Prev Res Phila Pa*. 2011 Sep;4(9):1401–8.
159. Gronlund B, Høgdall C, Hilden J, Engelholm SA, Høgdall EVS, Hansen HH. Should CA-125 response criteria be preferred to response evaluation criteria in solid tumors (RECIST) for prognostication during second-line chemotherapy of ovarian carcinoma? *J Clin Oncol Off J Am Soc Clin Oncol*. 2004 Oct 15;22(20):4051–8.
160. Rustin GJS, Quinn M, Thigpen T, du Bois A, Pujade-Lauraine E, Jakobsen A, et al. Re: New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). *J Natl Cancer Inst*. 2004 Mar 17;96(6):487–8.
161. Sturgeon CM, Duffy MJ, Hofmann BR, Lamerz R, Fritsche HA, Gaarenstroom K, et al. Laboratory Medicine Practice Guidelines for use of tumor markers in liver, bladder, cervical, and gastric cancers National Academy of Clinical Biochemistry. *Clin Chem*. 2010 Jun;56(6):e1–48.

162. Li D, Mallory T, Satomura S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin Chim Acta Int J Clin Chem*. 2001 Nov;313(1-2):15–9.
163. Sato Y, Nakata K, Kato Y, Shima M, Ishii N, Koji T, et al. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *N Engl J Med*. 1993 Jun 24;328(25):1802–6.
164. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Hepatobiliary Cancers [Internet]. 2014. Available from: [http://www.nccn.org/professionals/physician\\_gls/pdf/hepatobiliary.pdf](http://www.nccn.org/professionals/physician_gls/pdf/hepatobiliary.pdf)
165. Han X, Aslanian A, Yates JR. Mass spectrometry for proteomics. *Curr Opin Chem Biol*. 2008 Oct;12(5):483–90.
166. Peterman SM, Mulholland JJ. A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J Am Soc Mass Spectrom*. 2006 Feb;17(2):168–79.
167. Singh C, Zampronio CG, Creese AJ, Cooper HJ. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of *N*-Linked Glycoproteins. *J Proteome Res*. 2012 Sep 7;11(9):4517–25.
168. Alley WR, Mann BF, Novotny MV. High-sensitivity Analytical Approaches for the Structural Characterization of Glycoproteins. *Chem Rev*. 2013 Apr 10;113(4):2668–732.
169. Krokhin O, Ens W, Standing KG, Wilkins J, Perreault H. Site-specific *N*-glycosylation analysis: matrix-assisted laser desorption/ionization quadrupole-quadrupole time-of-flight tandem mass spectral signatures for recognition and identification of glycopeptides. *Rapid Commun Mass Spectrom RCM*. 2004;18(18):2020–30.
170. Thaysen-Andersen M, Packer NH. Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the *N*- and *O*-glycoproteome. *Biochim Biophys Acta BBA - Proteins Proteomics*. 2014 Sep;1844(9):1437–52.
171. Sullivan B, Addona TA, Carr SA. Selective detection of glycopeptides on ion trap mass spectrometers. *Anal Chem*. 2004 Jun 1;76(11):3112–8.
172. Sandra K, Devreese B, Van Beeumen J, Stals I, Claeysens M. The Q-Trap mass spectrometer, a novel tool in the study of protein glycosylation. *J Am Soc Mass Spectrom*. 2004 Mar;15(3):413–23.

173. Cao L, Tolić N, Qu Y, Meng D, Zhao R, Zhang Q, et al. Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem.* 2014 May;452:96–102.
174. Kubota K, Sato Y, Suzuki Y, Goto-Inoue N, Toda T, Suzuki M, et al. Analysis of glycopeptides using lectin affinity chromatography with MALDI-TOF mass spectrometry. *Anal Chem.* 2008 May 15;80(10):3693–8.
175. Leymarie N, Zaia J. Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal Chem.* 2012 Apr 3;84(7):3040–8.
176. Zhang Y, Jiao J, Yang P, Lu H. Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clin Proteomics.* 2014;11(1):18.
177. Lai ZW, Nice EC, Schilling O. Glyco-capture-based proteomics for secretome analysis. *PROTEOMICS.* 2013 Feb;13(3-4):512–25.
178. Pan S, Chen R, Aebersold R, Brentnall TA. Mass Spectrometry Based Glycoproteomics--From a Proteomics Perspective. *Mol Cell Proteomics.* 2011 Jan 1;10(1):R110.003251–R110.003251.
179. Kaji H, Yamauchi Y, Takahashi N, Isobe T. Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat Protoc.* 2007 Jan;1(6):3019–27.
180. Zhang H, Li X-J, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol.* 2003 Jun;21(6):660–6.
181. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, et al. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res.* 2009 Feb;8(2):651–61.
182. Berven FS, Ahmad R, Clauser KR, Carr SA. Optimizing performance of glycopeptide capture for plasma proteomics. *J Proteome Res.* 2010 Apr 5;9(4):1706–15.
183. Parker BL, Palmisano G, Edwards AVG, White MY, Engholm-Keller K, Lee A, et al. Quantitative N-linked Glycoproteomics of Myocardial Ischemia and Reperfusion Injury Reveals Early Remodeling in the Extracellular Environment. *Mol Cell Proteomics.* 2011 Aug 1;10(8):M110.006833–M110.006833.
184. Yeh C-H, Chen S-H, Li D-T, Lin H-P, Huang H-J, Chang C-I, et al. Magnetic bead-based hydrophilic interaction liquid chromatography for glycopeptide enrichments. *J Chromatogr A.* 2012 Feb;1224:70–8.



185. Mysling S, Palmisano G, Højrup P, Thaysen-Andersen M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem*. 2010 Jul 1;82(13):5598–609.
186. Alley WR, Mechref Y, Novotny MV. Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. *Rapid Commun Mass Spectrom*. 2009 Feb 28;23(4):495–505.
187. Engholm-Keller K, Larsen MR. Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds — Applications in acidic modification-specific proteomics. *J Proteomics*. 2011 Dec;75(2):317–28.
188. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NHH. Exploring the Sialome Using Titanium Dioxide Chromatography and Mass Spectrometry. *Mol Amp Cell Proteomics*. 2007 Jul 9;6(10):1778–87.
189. Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat Protoc*. 2010 Nov;5(12):1974–82.
190. Laughlin ST, Agard NJ, Baskin JM, Carrico IS, Chang PV, Ganguli AS, et al. Metabolic labeling of glycans with azido sugars for visualization and glycoproteomics. *Methods Enzymol*. 2006;415:230–50.
191. Hanson SR, Hsu T-L, Weerapana E, Kishikawa K, Simon GM, Cravatt BF, et al. Tailored Glycoproteomics and Glycan Site Mapping Using Saccharide-Selective Bioorthogonal Probes. *J Am Chem Soc*. 2007 Jun;129(23):7266–7.
192. Slade PG, Hajivandi M, Bartel CM, Gorfien SF. Identifying the CHO secretome using mucin-type O-linked glycosylation and click-chemistry. *J Proteome Res*. 2012 Dec 7;11(12):6175–86.
193. Yang L, Nyalwidhe JO, Guo S, Drake RR, Semmes OJ. Targeted Identification of Metastasis-associated Cell-surface Sialoglycoproteins in Prostate Cancer. *Mol Cell Proteomics*. 2011 Jun 1;10(6):M110.007294–M110.007294.
194. Roper SM, Zemsikova M, Neely BA, Martin A, Gao P, Jones EE, et al. Targeted glycoprotein enrichment and identification in stromal cell secretomes using azido sugar metabolic labeling. *PROTEOMICS - Clin Appl*. 2013 Jun;7(5-6):367–71.
195. Best MD. Click Chemistry and Bioorthogonal Reactions: Unprecedented Selectivity in the Labeling of Biological Molecules. *Biochemistry (Mosc)*. 2009 Jul 21;48(28):6571–84.

196. Frost DC, Li L. Recent Advances in Mass Spectrometry-Based Glycoproteomics. *Advances in Protein Chemistry and Structural Biology* [Internet]. Elsevier; 2014 [cited 2014 Nov 18]. p. 71–123. Available from: <http://linkinghub.elsevier.com/retrieve/pii/B9780128004531000038>
197. Gerken TA, Gupta R, Jentoft N. A novel approach for chemically deglycosylating O-linked glycoproteins. The deglycosylation of submaxillary and respiratory mucins. *Biochemistry (Mosc)*. 1992 Jan 28;31(3):639–48.
198. Edge ASB. Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochem J*. 2003 Dec 1;376(Pt 2):339–50.
199. Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR. Chemical Deamidation: A Common Pitfall in Large-Scale N-Linked Glycoproteomic Mass Spectrometry-Based Analyses. *J Proteome Res*. 2012 Mar 2;11(3):1949–57.
200. Zaia J. Mass Spectrometry and Glycomics. *OMICS J Integr Biol*. 2010 Aug;14(4):401–18.
201. Baycin Hizal D, Wolozny D, Colao J, Jacobson E, Tian Y, Krag SS, et al. Glycoproteomic and glycomic databases. *Clin Proteomics*. 2014;11(1):15.
202. Bern M, Kil YJ, Becker C. Byonic: Advanced Peptide and Protein Identification Software. In: Baxevanis AD, Petsko GA, Stein LD, Stormo GD, editors. *Current Protocols in Bioinformatics* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2012 [cited 2014 Dec 1]. Available from: <http://doi.wiley.com/10.1002/0471250953.bi1320s40>
203. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B*. 2007 Apr;849(1-2):115–28.
204. Segu ZM, Mechref Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom*. 2010 May 15;24(9):1217–25.
205. Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AVG, Crossett B, et al. Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of *Campylobacter jejuni*. *Mol Cell Proteomics*. 2011 Feb 1;10(2):M000031–MCP201 – M000031–MCP201.

206. Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R. Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology*. 2006 May;16(5):82R – 90R.
207. Aoki-Kinoshita KF. Using databases and web resources for glycomics research. *Mol Cell Proteomics MCP*. 2013 Apr;12(4):1036–45.
208. Click-iT Protein Enrichment Kit [Internet]. Invitrogen; 2011. Available from: <https://tools.lifetechnologies.com/content/sfs/manuals/mp10416.pdf>
209. Urbaniak MD, Martin DMA, Ferguson MAJ. Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*. *J Proteome Res*. 2013 May 3;12(5):2233–44.
210. Quackenbush J. Microarray data normalization and transformation. *Nat Genet*. 2002 Dec;32 Suppl:496–501.
211. Powers TW, Jones EE, Betesh LR, Romano PR, Gao P, Copland JA, et al. Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked glycan expression in tissues. *Anal Chem*. 2013 Oct 15;85(20):9799–806.
212. Reiding KR, Blank D, Kuijper DM, Deelder AM, Wührer M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem*. 2014 Jun 17;86(12):5784–93.
213. Mechref Y, Kang P, Novotny MV. Solid-phase permethylation for glycomic analysis. *Methods Mol Biol Clifton NJ*. 2009;534:53–64.
214. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans<sup>†</sup>. *J Proteome Res*. 2008 Apr;7(4):1650–9.
215. Orsburn BC. SILAC in Biomarker Discovery. In: Zhou M, Veenstra T, editors. *Proteomics for Biomarker Discovery* [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2015 Feb 4]. p. 123–31. Available from: [http://link.springer.com/10.1007/978-1-62703-360-2\\_11](http://link.springer.com/10.1007/978-1-62703-360-2_11)
216. Vagin O, Tokhtaeva E, Sachs G. The role of the beta1 subunit of the Na,K-ATPase and its glycosylation in cell-cell adhesion. *J Biol Chem*. 2006 Dec 22;281(51):39573–87.
217. Bab-Dinitz E, Albeck S, Peleg Y, Brumfeld V, Gottschalk KE, Karlish SJD. A C-terminal lobe of the beta subunit of Na,K-ATPase and H,K-ATPase resembles cell adhesion molecules. *Biochemistry (Mosc)*. 2009 Sep 15;48(36):8684–91.

218. Velonas V, Woo H, Remedios C, Assinder S. Current Status of Biomarkers for Prostate Cancer. *Int J Mol Sci*. 2013 May 24;14(6):11034–60.
219. Turtoi A, Dumont B, Greffe Y, Blomme A, Mazzucchelli G, Delvenne P, et al. Novel comprehensive approach for accessible biomarker identification and absolute quantification from precious human tissues. *J Proteome Res*. 2011 Jul 1;10(7):3160–82.
220. Azuma K, Serada S, Takamatsu S, Terao N, Takeishi S, Kamada Y, et al. Identification of Sialylated Glycoproteins in Doxorubicin-Treated Hepatoma Cells with Glycoproteomic Analyses. *J Proteome Res*. 2014 Nov 7;13(11):4869–77.
221. Kim J, Myers AC, Chen L, Pardoll DM, Truong-Tran Q-A, Lane AP, et al. Constitutive and Inducible Expression of B7 Family of Ligands by Human Airway Epithelial Cells. *Am J Respir Cell Mol Biol*. 2005 Sep;33(3):280–9.
222. Xue. The distribution and expression profiles of human Aspartyl/Asparaginyl beta-hydroxylase in tumor cell lines and human tissues. *Oncol Rep [Internet]*. 2010 Sep 27 [cited 2015 Feb 3];24(5). Available from: <http://www.spandidos-publications.com/or/24/5/1257>
223. Bateman A, Bennett HPJ. The granulin gene family: from cancer to dementia. *BioEssays News Rev Mol Cell Dev Biol*. 2009 Nov;31(11):1245–54.
224. He Z, Ismail A, Kriazhev L, Sadvakassova G, Bateman A. Progranulin (PC-cell-derived growth factor/acrogranin) regulates invasion and cell survival. *Cancer Res*. 2002 Oct 1;62(19):5590–6.
225. Songsrirote K, Li Z, Ashford D, Bateman A, Thomas-Oates J. Development and application of mass spectrometric methods for the analysis of progranulin N-glycosylation. *J Proteomics*. 2010 Jun 16;73(8):1479–90.
226. Lewandrowski U, Moebius J, Walter U, Sickmann A. Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach. *Mol Cell Proteomics MCP*. 2006 Feb;5(2):226–33.
227. Yang H, Zubarev RA. Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides: General. *ELECTROPHORESIS*. 2010 Jun;31(11):1764–72.
228. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol*. 2003 Jun;21(6):667–72.

229. Park Y, Lebrilla CB. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrom Rev.* 2005 Mar;24(2):232–64.
230. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol.* 2006 Aug;24(8):971–83.
231. Nguyen EV, Gharib SA, Schnapp LM, Goodlett DR. Shotgun MS proteomic analysis of bronchoalveolar lavage fluid in normal subjects. *Proteomics Clin Appl.* 2014 Oct;8(9-10):737–47.
232. Magi B, Bargagli E, Bini L, Rottoli P. Proteome analysis of bronchoalveolar lavage in lung diseases. *PROTEOMICS.* 2006 Dec;6(23):6354–69.
233. Meyer KC. Bronchoalveolar lavage as a diagnostic tool. *Semin Respir Crit Care Med.* 2007 Oct;28(5):546–60.
234. Balog CIA, Stavenhagen K, Fung WLJ, Koeleman CA, McDonnell LA, Verhoeven A, et al. N-glycosylation of Colorectal Cancer Tissues: A liquid chromatography and mass spectrometry-based investigation. *Mol Cell Proteomics.* 2012 Sep 1;11(9):571–85.
235. Caprioli RM, Farmer TB, Gile J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal Chem.* 1997 Dec;69(23):4751–60.
236. Jones EE, Powers TW, Neely BA, Cazares LH, Troyer DA, Parker AS, et al. MALDI imaging mass spectrometry profiling of proteins and lipids in clear cell renal cell carcinoma. *PROTEOMICS.* 2014 Apr;14(7-8):924–35.
237. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GENE SeT ANALYSIS Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 2013 Jul;41(Web Server issue):W77–83.
238. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W741–8.
239. Powers TW, Neely BA, Shao Y, Tang H, Troyer DA, Mehta AS, et al. MALDI Imaging Mass Spectrometry Profiling of N-Glycans in Formalin-Fixed Paraffin Embedded Clinical Tissue Blocks and Tissue Microarrays. *Batra SK, editor. PLoS ONE.* 2014 Sep 3;9(9):e106255.
240. Foster MW, Morrison LD, Todd JL, Snyder LD, Thompson JW, Soderblom EJ, et al. Quantitative proteomics of bronchoalveolar lavage fluid in idiopathic pulmonary fibrosis. *J Proteome Res.* 2015 Feb 6;14(2):1238–49.

241. Zaia J. Mass spectrometry of oligosaccharides. *Mass Spectrom Rev.* 2004 Jun;23(3):161–227.
242. Lenz AG, Meyer B, Costabel U, Maier K. Bronchoalveolar lavage fluid proteins in human lung disease: analysis by two-dimensional electrophoresis. *Electrophoresis.* 1993 Mar;14(3):242–4.
243. De Torre C, Ying S-X, Munson PJ, Meduri GU, Suffredini AF. Proteomic analysis of inflammatory biomarkers in bronchoalveolar lavage. *Proteomics.* 2006 Jul;6(13):3949–57.
244. Drake RR, Cazare LH, Semmes OJ, Wadsworth JT. Serum, salivary and tissue proteomics for discovery of biomarkers for head and neck cancers. *Expert Rev Mol Diagn.* 2005 Jan;5(1):93–100.
245. Hu S, Arellano M, Boontheung P, Wang J, Zhou H, Jiang J, et al. Salivary proteomics for oral cancer biomarker discovery. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2008 Oct 1;14(19):6246–52.
246. Vieira AC, An HJ, Ozcan S, Kim J-H, Lebrilla CB, Mannis MJ. Glycomic analysis of tear and saliva in ocular rosacea patients: the search for a biomarker. *Ocul Surf.* 2012 Jul;10(3):184–92.
247. Xiao H, Zhang L, Zhou H, Lee JM, Garon EB, Wong DTW. Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Mol Cell Proteomics MCP.* 2012 Feb;11(2):M111.012112.
248. Oztürk LK, Emekli-Alturfan E, Kaşıkci E, Demir G, Yarat A. Salivary total sialic acid levels increase in breast cancer patients: a preliminary study. *Med Chem Shāriqah United Arab Emir.* 2011 Sep;7(5):443–7.
249. Ruhl S. The scientific exploration of saliva in the post-proteomic era: from database back to basic function. *Expert Rev Proteomics.* 2012 Feb;9(1):85–96.
250. Fumagalli M, Ferrari F, Luisetti M, Stolk J, Hiemstra P, Capuano D, et al. Profiling the Proteome of Exhaled Breath Condensate in Healthy Smokers and COPD Patients by LC-MS/MS. *Int J Mol Sci.* 2012 Oct 29;13(12):13894–910.
251. Soltermann A, Ossola R, Kilgus-Hawelski S, von Eckardstein A, Suter T, Aebersold R, et al. N-glycoprotein profiling of lung adenocarcinoma pleural effusions by shotgun proteomics. *Cancer.* 2008 Mar 7;114(2):124–33.
252. Jurchen JC, Rubakhin SS, Sweedler JV. MALDI-MS imaging of features smaller than the size of the laser beam. *J Am Soc Mass Spectrom.* 2005 Oct;16(10):1654–9.

253. Zavalin A, Todd EM, Rawhouser PD, Yang J, Norris JL, Caprioli RM. Direct imaging of single cells and tissue at sub-cellular spatial resolution using transmission geometry MALDI MS. *J Mass Spectrom.* 2012 Nov;47(11):i – i.
254. Bremnes RM, Al-Shibli K, Donnem T, Sirera R, Al-Saad S, Andersen S, et al. The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer.* 2011 Apr;6(4):824–33.
255. Suzuki K, Kachala SS, Kadota K, Shen R, Mo Q, Beer DG, et al. Prognostic Immune Markers in Non-Small Cell Lung Cancer. *Clin Cancer Res.* 2011 Aug 15;17(16):5247–56.
256. Principe S, Jones EE, Kim Y, Sinha A, Nyalwidhe JO, Brooks J, et al. In-depth proteomic analyses of exosomes isolated from expressed prostatic secretions in urine. *PROTEOMICS.* 2013 May;13(10-11):1667–71.
257. Duijvesz D, Luidert T, Bangma CH, Jenster G. Exosomes as biomarker treasure chests for prostate cancer. *Eur Urol.* 2011 May;59(5):823–31.
258. Fang M, Lim J-M, Wells L. Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG). In: Arkin AP, Mahal L, Romesberg F, Shah K, Shamu C, Thomas C, editors. *Current Protocols in Chemical Biology* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010 [cited 2015 Feb 20]. Available from: <http://doi.wiley.com/10.1002/9780470559277.ch090207>
259. Orlando R, Lim J-M, Atwood JA, Angel PM, Fang M, Aoki K, et al. IDAWG: Metabolic Incorporation of Stable Isotope Labels for Quantitative Glycomics of Cultured Cells. *J Proteome Res.* 2009 Aug 7;8(8):3816–23.
260. Chen C-D, Wang C-L, Yu C-J, Chien K-Y, Chen Y-T, Chen M-C, et al. Targeted Proteomics Pipeline Reveals Potential Biomarkers for the Diagnosis of Metastatic Lung Cancer in Pleural Effusion. *J Proteome Res.* 2014 Jun 6;13(6):2818–29.
261. Conrad DH, Goyette J, Thomas PS. Proteomics as a Method for Early Detection of Cancer: A Review of Proteomics, Exhaled Breath Condensate, and Lung Cancer Screening. *J Gen Intern Med.* 2008 Jan;23(S1):78–84.
262. Norris JL, Caprioli RM. Imaging mass spectrometry: A new tool for pathology in a molecular age. *PROTEOMICS - Clin Appl.* 2013 Dec;7(11-12):733–8.
263. Korekane H, Shida K, Murata K, Ohue M, Sasaki Y, Imaoka S, et al. Evaluation of laser microdissection as a tool in cancer glycomic studies. *Biochem Biophys Res Commun.* 2007 Jan 19;352(3):579–86.

## Biography

Stephen Roper was born in Kerrville, TX. in December of 1979. He is the youngest of four children, including two sisters and a brother. His parents, John and Evelyn Roper, worked as school teachers and always encouraged Stephen to pursue his curiosity in the medical sciences. After graduating from high school in 1998, Stephen attended junior college where he majored in vocational nursing and obtained his license to practice as a nurse in the state of Texas. He continued his education at Texas Tech University where he received a Bachelor's of Science degree in Biology in 2004. He worked as a North Pacific Groundfish Observer in the Bering Sea for a year before he decided to return to school for further training in the biological sciences. In 2007, Stephen graduated from Texas Tech University Health Sciences Center with a Master's of Science degree in Molecular Pathology. He obtained employment as a medical technologist at the University of Texas MD Anderson Cancer Center in the molecular diagnostics lab. After two years of service in this setting, Stephen decided that he would like to continue his education by pursuing a PhD in Biomedical Science. He accepted a position in the Eastern Virginia Medical School graduate program in 2009 and began his studies in cancer proteomics under the direction of Dr. Richard Drake. In July 2011, Dr. Drake was offered an esteemed position at the Medical University of South Carolina (MUSC) and invited Stephen to continue his studies with him at this new school. After four additional years of study at MUSC, Stephen obtained his PhD in Biomedical Science in May of 2015.