

Medical University of South Carolina

MEDICA

MUSC Theses and Dissertations

Spring 4-10-2023

Utilizing Mass Spectrometry Imaging to Correlate N-Glycosylation of Hepatocellular Carcinoma with Tumor Subtypes for Biomarker Discovery

Andrew DelaCourt

Medical University of South Carolina

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



Part of the [Cancer Biology Commons](#)

Recommended Citation

DelaCourt, Andrew, "Utilizing Mass Spectrometry Imaging to Correlate N-Glycosylation of Hepatocellular Carcinoma with Tumor Subtypes for Biomarker Discovery" (2023). *MUSC Theses and Dissertations*. 775. <https://medica-musc.researchcommons.org/theses/775>

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.

Utilizing Mass Spectrometry Imaging to Correlate N-Glycosylation of
Hepatocellular Carcinoma with Tumor Subtypes for Biomarker Discovery

by

Andrew Taylor DelaCourt

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

Department of Cell and Molecular Pharmacology and Experimental Therapeutics

2023

Approved by:

Chairman, Advisory Committee

Anand Mehta

Peggi Angel

Nathan Dolloff

Stephen Duncan

Yuri Peterson

TABLE OF CONTENTS

ABSTRACT.....	vii
LIST OF FIGURES.....	ix
LIST OF TABLES.....	x
ACKNOWLEDGEMENTS.....	xi
Chapter 1: Introduction.....	1
1.1 Introduction to Hepatocellular Carcinoma.....	3
1.1.1 Disease Etiology, Development, and Progression.....	3
1.1.2 Genetics and Molecular Pathways.....	7
1.1.3 Treatment Options and Outcomes.....	12
1.1.4 Surveillance and Biomarkers.....	20
1.2 N-linked Glycosylation.....	26
1.2.1 Broad Overview of Glycosylation.....	26
1.2.2 Synthesis and Structure of N-Glycans.....	30
1.2.3 Dysregulated N-Glycosylation in Cancer.....	33
1.2.4 N-Glycosylation in Cancer Diagnostics.....	38
1.2.5 Glycomics Methods of N-Glycan Analysis.....	44
1.2.5.1 Biosynthetic Inhibitors of N-Glycans.....	44
1.2.5.2 Lectin-based Analysis of N-Glycans.....	45
1.2.5.3 High Performance Liquid Chromatography of N-Glycans.....	47
1.2.5.4 Mass Spectrometry-based Analysis of N-Glycans.....	49
1.3 Overview of Mass Spectrometry.....	50
1.3.1 Basic Principles of Mass Spectrometry.....	50
1.3.2 Mass Spectrometry Imaging.....	53
1.3.3 Mass Spectrometers in this Dissertation.....	55
1.4 Broad Overview.....	57
Chapter 2: Hypothesis.....	60
2.1 Rationale.....	61
2.2 Significance.....	62
2.3 Innovation.....	63
2.4 Specific Aim 1.....	64
2.5 Specific Aim 2.....	65

Chapter 3: Tissue N-Glycosylation Correlates with Hepatocellular Carcinoma	
Subtypes	67
3.1 Introduction.....	69
3.2 Materials and Methods.....	72
3.2.1 Patient Tissues and Microarrays.....	72
3.2.2 FFPE Tissue Preparation for MALDI-IMS.....	73
3.2.3 MALDI-IMS of N-Glycans.....	76
3.2.4 Statistical Analysis.....	76
3.3 Results.....	78
3.3.1 Inter-Tumor Heterogeneity of N-Glycan Expression in HCC Tissues.....	78
3.3.2 Tumor-Specific N-Glycan Expression Associated with HCC Molecular Subtypes.....	85
3.3.3 Validation of Differentiation of Subtypes by N-Glycosylation through MALDI-TOF.....	93
3.4 Discussion.....	98
Chapter 4: Novel Enzymatic Approach to Analyze De-Sialylated N-linked Glycans	103
4.1 Introduction.....	105
4.2 Materials and Methods.....	108
4.2.1 Cloning, Expression, and Purification of a Recombinant Sialidase	108
4.2.2 In-Solution Digestion by Recombinant Sialidase.....	109
4.2.3 N-Glycan Sequencing.....	109
4.2.4 Patient Tissues.....	110
4.2.5 FFPE Tissue Preparation for MALDI-IMS.....	110
4.2.6 N-Glycan Imaging via MALDI-IMS.....	111
4.2.7 UPLC of Tissue Extracted N-Glycans.....	115
4.3 Results.....	116
4.3.1 In-Solution Analysis of Sialidase Prime™ Activity on N-Glycans.....	116
4.3.2 On-Tissue Analysis of Sialidase Activity using MALDI-IMS.....	120
4.3.3 UPLC Validation of On-Tissue Sialidase Activity.....	127
4.4 Discussion.....	131
Chapter 5: N-Glycomic Analysis of Patient-Matching Serum and Tissue	134
5.1 Introduction.....	135
5.2 Materials and Methods	138
5.2.1 Patient Tissues and Serum.....	138
5.2.2 FFPE Tissue Preparation.....	138
5.2.3 N-Glycomic Total Serum Preparation.....	139

5.2.4 Serum Glycoprotein Antibody Array Preparation.....	139
5.2.5 MALDI-IMS N-Glycan Imaging.....	140
5.2.6 Statistical Data Analysis.....	143
5.3 Results.....	144
5.3.1 Demographic and Clinical Information of Sample Cohort.....	144
5.3.2 N-Glycosylation of HCC Compared to Background Liver.....	148
5.3.3 Distinguishing HCC Subtypes in Tissue.....	152
5.3.4 Comparison of Tissue and Serum N-Glycosylation.....	158
5.3.5 Subtyped Serum Analysis.....	160
5.4 Discussion.....	166
Chapter 6: Conclusions, Limitations, and Future Studies.....	170
6.1 Overall Findings.....	171
6.2 Analysis of the N-Glycosylation of HCC Tumor Subtypes.....	171
6.2.1 Conclusions.....	171
6.2.2 Limitations and Future Studies.....	172
6.3 Matching Serum and Tissue N-Glycan Analysis for HCC Biomarker Development.....	173
6.3.1 Conclusions.....	173
6.3.2 Limitations and Future Studies.....	174
6.4 Utilization of Sialidase for MALDI-IMS N-Glycan Imaging.....	175
6.4.1 Conclusions.....	175
6.4.2 Limitations and Future Studies.....	175
6.5 Final Thoughts.....	176
REFERENCES.....	177

ABSTRACT

ANDREW TAYLOR DELACOURT. Utilizing Mass Spectrometry Imaging to Correlate N-Glycosylation of Hepatocellular Carcinoma with Tumor Subtypes for Biomarker Discovery. (Under the direction of ANAND MEHTA)

Hepatocellular carcinoma (HCC) is a leading cause of cancer deaths globally and is a growing clinical problem with poor survival outcomes beyond early-stage disease. Surveillance for HCC has primarily relied on ultrasound and serum α -fetoprotein (AFP), but combined they only have a sensitivity of 63% for early-stage HCC tumors, suggesting a need for improved diagnostic strategies. Alterations to N-glycan expression are relevant to the progression of cancer, and there a multitude of N-glycan-based cancer biomarkers that have been identified with sensitivity for various cancer types including HCC. Spatial HCC tissue profiling of N-linked glycosylation by matrix-assisted laser desorption ionization imaging mass spectrometry (MALDI-IMS) serves as a new method to evaluate tumor-correlated N-glycosylation and thereby identify potential HCC biomarkers. Previous work has identified significant changes in the N-linked glycosylation of HCC tumors, but has not accounted for the heterogeneous genetic and molecular nature of HCC, which has led to inadequate sensitivity of N-glycan biomarkers. Therefore, we hypothesized that the incorporation of genetic/molecular information into N-glycan-based biomarker development would result in improved sensitivity for HCC. To determine the correlation between HCC-specific N-glycosylation

and genetic/molecular tumor features, we profiled HCC tissue samples with MALDI-IMS and correlated the spatial N-glycosylation with a widely used HCC molecular classification that utilizes histological, genetic, and clinical tumor features (Hoshida subtypes). MALDI-IMS data displayed trends that could approximately distinguish between subtypes, with Subtype 1 demonstrating significantly dysregulated N-glycosylation compared to Subtypes 2 and 3, particularly in regard to fucosylation. In order to further the clinical relevance of subtype-dependent N-glycosylation, we analyzed patient-matching HCC tumor tissue, background liver tissue and serum samples through MALDI-IMS. Results showed a N-glycan based model capable of differentiating tumor tissue from background liver tissue with an AUC of 0.9842. When analyzing the associated serum, 24.7% of detected N-glycans were significantly positively correlated between tumor tissue and serum, suggesting that N-glycosylation trends translate from tissue to serum. Additionally, a serum N-glycan-based model was capable of distinguishing Subtype 1/Subtype 2 tumors from Subtype 3 tumors with an AUC of 0.881. Through the utilization of MALDI-IMS, subtype-dependent N-glycosylation trends were identified in both tissue and serum, which can significantly further the development of HCC biomarkers for clinical application.

LIST OF FIGURES

Chapter 1:

Figure 1. Hoshida HCC Subtyping System.....	11
Figure 2. Hepatocellular Carcinoma Staging and Outcomes.....	14
Figure 3. Monosaccharide Structural Representations.....	27
Figure 4. N-Glycan Structural Classifications.....	31

Chapter 3:

Figure 5. MALDI-IMS Tissue Preparation Workflow.....	75
Figure 6. TMA Analysis of the N-glycosylation of HCC and Normal Liver.....	80
Figure 7. N-Glycosylation of HCC Subtypes.....	87
Figure 8. Representative S1 Tumors.....	90
Figure 9. Fucosylation of Subtype 2 Tumors.....	92
Figure 10. MALDI-TOF Validation Cohort.....	97

Chapter 4:

Figure 11. Sialidase Prime Activity Validation.....	117
Figure 12. Sialidase Prime Activity on Cleaved N-Glycans.....	119
Figure 13. Tissue H&E Stains.....	122
Figure 14. MALDI-IMS with Simultaneous PNGase F and Sialidase Incubation.....	123
Figure 15. MALDI-IMS with Simultaneous Endo F3 and Sialidase Incubation.....	126
Figure 16. UPLC Analysis of On-Tissue Sialidase/PNGase F Activity.....	129
Figure 17. Analysis of Sialylated Glycoforms via MALDI-IMS.....	130

Chapter 5:

Figure 18. Example m/z Images of Tumor and Background Liver Tissue.....	149
Figure 19. Classification Model for Tumor and Background Liver Tissue.....	151
Figure 20. Best Individual N-glycans for Subtype Discrimination in Tumor and Background Liver.....	153
Figure 21. N-Glycans to Distinguish Each Subtype Pairwise.....	155
Figure 22. Optimized Model to Classify HCC Subtypes Utilizing N-Glycosylation.....	157
Figure 23. Pearson Correlations of Tumor and Serum N-Glycans.....	159
Figure 24. Total Serum N-Glycan Model to Differentiate Subtypes.....	161
Figure 25. Antibody Array Experimental Set Up.....	163
Figure 26. Glycoprotein-Specific Model to Differentiate Subtypes.....	165

LIST OF TABLES

Chapter 3:

Table 1. Patient Demographics.....	73
Table 2. N-Glycans Analyzed on the MALDI FT-ICR.....	82
Table 3. Significantly HCC-associated N-glycans within TMA.....	83
Table 4. Significantly Normal-Associated N-glycans within TMA.....	84
Table 5. N-Glycans analyzed on the MALDI-TOF.....	95

Chapter 4:

Table 6. N-Glycan Peak List with PNGase F.....	113
Table 7. N-Glycan Peak List with Endo F3.....	114
Table 8. Signal Intensity with PNGase F and with PNGase F/Sialidase.....	124

Chapter 5:

Table 9. Combined N-Glycan Peak List for Tissue/Serum.....	142
Table 10. Demographic/Clinical Information of Tissue Sample Cohort.....	145
Table 11. Demographic/Clinical Information of Serum Sample Cohort.....	147

ACKNOWLEDGEMENTS

I have so many people that I'd like to thank who have helped me reach the accomplishment of writing and defending this dissertation. To start, I'd like to sincerely thank my mentor Dr. Mehta for his unwavering support throughout the entirety of my Ph.D. education. I truly believe that his mentorship has been critical for me to become the scientist that I am today, and I am extremely grateful for the manner in which he has supported my scientific ideas while challenging me to push the limits of what I thought was possible. I'd also like to acknowledge the entirety of my thesis committee, who have all been extremely supportive throughout my time at MUSC. Dr. Angel, thank you for teaching me so much about mass spectrometry imaging and inspiring me to continue finding new and exciting applications for the technology. Dr. Dolloff, Dr. Peterson, and Dr. Duncan, thank you for serving on my committee and always supporting my development as a scientist. I am greatly appreciative of all of your voices and how you all have challenged me to think not only about mass spectrometry, but also about the important biological and clinical aspects of my research. Additionally, I want to thank Dr. Drake, who I consider as an unofficial committee member. I am deeply grateful for all of your support in the lab throughout the years, you have created a great environment for graduate students to thrive at MUSC within the areas of mass spectrometry and glycobiology.

There are so many students at MUSC who I owe gratitude for their support and camaraderie. Calvin, you've always been there for me to talk both the good and the bad things through, and I'm deeply grateful for you as both a peer and a friend. Colin,

between living together and working together you were an integral part of my experience at MUSC, and you were a great friend who made going through graduate school a much better experience. To both of you, thanks for all the memories from being stranded on a layover in Detroit, to spending entire days fixing the sprayer, to challenging each other every morning in geoguesser. I'd also like to extend a sincere thanks to all students that have come through the Mehta lab during my time (Alyson, Sharon, Jake, Jake, and Harmin), who each have contributed meaningfully to my experience. Alyson, thanks for mentoring me as a rotation student and for always being there to support me, graduate school would have been way different without your positivity (and without your desk). Sharon, thanks for coming back to the Mehta lab, you've been a foundational element for me throughout our time in the lab. Jake, thanks for always bringing your tough scientific questions to me, you've always challenged me to think critically about the science that we do and I'm a better scientist because of it. Jake, thanks for bringing a new element to our lab with your project. Additionally, I'd like to thank all of the other students in the adjacent mass spectrometry labs (Danielle, Connor, Casey, Denys, Jordan, Jadey, Jaclyn, Rachel, Elizabeth, and Stephen), as each one of you has contributed to making our labs such an enjoyable place to come in to work every day.

Finally, I want to thank my family and friends who have supported me throughout graduate school and have allowed me to reach this accomplishment. To my parents, thank you for never wavering in your support of my dreams, and for always being there whenever I needed anything. To my siblings Bryce and Sarah, thanks for

helping me to become the man that I am today and for inspiring me with your own accomplishments. There are too many friends both from Wake Forest and Charleston to name here, but just know that I am grateful for each and every one of you. To my girlfriend Grace, thank you for always supporting me through the end of graduate school, your constant encouragement is a big reason why I have made it to this point. Lastly, I'd like to thank God for paving the way for me to reach this point in my career, I certainly would not have made it here without His faithfulness, and I am excited to see the plans He has for me in my upcoming professional scientific career.

Chapter 1: Introduction

Select sections of this chapter have been adapted from two separate manuscripts. The first is published in *Comprehensive Pharmacology*, 2021: 112-125. ATD wrote the full manuscript, with editing and intellectual contributions from ASM. The second is published in *Advances in Cancer Research*, 2023 (157): 57-81. ATD wrote the full manuscript, with editing and intellectual contributions from ASM.

Liver Cancer (Current Therapies)

Andrew DelaCourt, and Anand S. Mehta

Beyond glyco-proteomics— Understanding the role of genetics in cancer biomarkers

Andrew DelaCourt, and Anand S. Mehta

1.1 Introduction to Hepatocellular Carcinoma

1.1.1 Disease Etiology, Development, and Progression

Liver cancer is the fifth most common cancer in men, the ninth most common in women, and the third leading cause of cancer deaths globally.¹⁻³ The most common form of liver cancer is Hepatocellular carcinoma (HCC), and over the past several decades HCC has been the fastest rising cause of cancer deaths in the United States due to a rising incidence and mortality rate.⁴ The five-year survival rate of HCC is only 18%, which in large part is due to the 60-70% of tumors that are diagnosed at late stages and are ineligible for curative surgical treatments.⁵ HCC primarily develops in the background of liver cirrhosis, which is an inflammatory response to chronic liver injury characterized by regenerative nodules and fibrosis.^{6,7} Risk factors include viral infections, alcohol consumption, obesity, metabolic disorders, and smoking.^{8,9} Viral infections most associated with HCC include Hepatitis B and C (HBV/HCV), which cause significant liver damage and are historically are the largest risk factors for HCC.¹⁰⁻¹² Based on these risk factors, there is a clearly defined population for screening of cirrhosis and cancer, which has shown demonstrated benefit to early stage detection and overall survival.¹³ Screening is exceedingly important to early detection as HCC does not present with symptoms until late-stage disease when survival outcomes are exceedingly poor.^{14,15}

The connection between HBV/HCV and HCC development has been extremely well validated, although the two viruses have varying disease populations. HBV is endemic in many eastern Asian and African countries, where vertical transmission is

frequent.⁹ There has been a demonstrated increase in HCC incidence with HBV cases with chronic cirrhosis when compared to HBV cases without cirrhosis or inactive carriers of HBV.¹⁶ HBV viral load has been demonstrated to correlate with HCC risk, even when adjusting for potential confounding factors such as alcohol use and smoking.¹⁷ HBV is divided into 8 genotypes, A-H, which differ based on the presence of mutations in the promotor and pre-core regions, which affect viral activity and thereby HCC risk.¹⁸ The most common HBV genotypes in eastern Asia, are genotypes B and C. HBV genotype C has been shown to be the most correlated with HCC development, and has also been shown to be more prevalent with increased age.^{19,20} Meanwhile, genotype B has been observed to have a higher prevalence in younger age groups and exhibited decreased cirrhosis.²¹ In addition to the promotion of cirrhosis development, HBV is thought to induce cancer-promoting mutations through viral replication, which is supported by evidence that inactive HBV still increases HCC risk through viral integration.²² Beyond liver injury generally, the specific molecular events that cause HBV infections to promote HCC are still not fully understood, and efforts are more concentrated on treating and preventing HBV itself through vaccination, which is an effective way to decrease HCC incidence indirectly.

HCV infections are geographically spread very differently than HBV, with concentrated areas in Europe, the United States, and Japan, where it has spread primarily through drug use. An increase in HCV infections in the United States has played a role in the increase in HCC incidence in recent years, but HCV infections are expected to level off in coming years and only partially explain the rising incidence rate

of HCC. There is significant evidence that HCV infections can cause HCC tumors, with data suggesting that a 15-20 fold increase in incidence occurs in HCV-positive subjects compared to HCV-negative subjects.^{9,23} Unlike HBV, there has not been an established association between HCV viral load and progression to cirrhosis or HCC.²⁴ Although less established than the link between HBV genotypes and HCC, reports have suggested that HCV genotype 1b has increased HCC risk compared to other HCV genotypes.²⁵ For both HBV and HCV, men have significantly higher HCC risk than women, which has been linked to hormonal differences acting as tumor promoters/suppressors.^{26,27}

There is substantial evidence that excessive alcohol use leads to HCC, although there is little to no evidence that low or moderate levels of alcohol usage has any effect on HCC incidence.²⁸ Even still, the number of individuals at risk for HCC due to alcohol abuse is substantial, particularly when that lifestyle is combined with additional risk factors. Alcohol usage has been demonstrated to have an additive effect with many conditions towards the development of HCC, particularly HBV and HCV.²⁸ The mechanism by which alcohol may act as a cancer promoting agent, outside of the obvious promotion of cirrhosis through liver injury, is not fully known. Alcohol is metabolized by cytochrome P450 in the liver, and it is possible that metabolites have cancer promoting activity and at the least are responsible for liver damage that fosters cirrhosis development over time. In particular, acetaldehyde and reactive oxygen species from the liver's breakdown of ingested alcohol may cause DNA damage directly to hepatocytes in order to promote HCC initiation.²⁹ While many of the specific molecular events are still not fully understood, it is clear that alcohol, similarly to HBV

and HCV, promotes liver injury both directly and indirectly, which can lead to HCC development in the long term.

Risk factors related to metabolic syndrome, including obesity and diabetes, are becoming very widespread, and even though affected patients develop HCC at lower rates, the large number of affected individuals make these risk factors the primary cause for increasing HCC incidence. In areas with low levels of viral hepatitis, HCC related mortality and incidence is increasing, which is expected to be explained by increasing rates of metabolic syndrome and related conditions.³⁰ Nonalcoholic fatty liver disease (NAFLD) is the leading cause of liver disease in the United States, and can lead to nonalcoholic steatohepatitis (NASH) and ultimately HCC. Around 20% of NAFLD cases eventually present as NASH, which features inflammation, liver injury, and a risk of progressing to cirrhosis and to HCC.³¹ There is also increasing evidence that NAFLD cases allow for the development of HCC in the absence of cirrhosis at a high rate, with one study suggesting that only 46% of NAFLD-related HCC cases had evidence of cirrhosis.³² NAFLD prevalence has been shown to be rapidly increasing in the United States in recent years, particularly among younger people.³³ In addition to NAFLD, diabetes has been shown to lead to increased HCC risk independent of obesity, and rates of obesity have been shown to be correlated with fatty liver disease and high HCC mortality rates.^{34,35} The prevalence of obesity and other metabolic risk factors leads to a frequent overlap with additional risk factors, such as alcohol usage and hepatitis infections, which accelerates HCC development.

1.1.2 Genetics and Molecular Pathways

There are not universal molecular events or oncogenes/tumor suppressor genes that lead to HCC initiation, which follows from the variety of risk factors and liver conditions that can lead to HCC development. As previously stated, cirrhosis precedes HCC in the majority of cases, particularly those developed from viral hepatitis and alcohol abuse. Cirrhosis is a condition in which hepatocytes are killed and an overabundance of connective tissue and inflammatory cells appears, which often leads to the development and proliferation of dysplastic hepatocytes.³⁶ Simplistically, this provides an environment filled with cell injury and inflammation, which necessitates proliferation of hepatocytes to maintain liver function and thus provides an opportunity for tumor-promoting mutations to occur. Tumor initiation of HCC has been demonstratively shown to be a very heterogenous event, with a significantly large number of genes and pathways with activity affected in some way. Some of these affected pathways and molecular events are impacted directly instead of indirectly through cirrhosis, and this is particularly true in the case of obesity-related HCC. Metabolic disorders put strain on the liver for a variety of molecular reasons, including insulin resistance, adipose-driven inflammation, and lipotoxicity, which can be expected to promote oncogenic pathways both in the presence and absence of cirrhosis.³¹ All of these pathways are known to promote obesity-related liver disease, and are capable of promoting liver damage and HCC growth even in the absence of cirrhosis.

Analyses of genomic sequencing of HCC cases have shown that there are several common mutations that may act as cancer drivers in HCC, including mutations to the

telomeres reverse transcriptase (TERT) promotor region, the common tumor suppressor gene *TP53*, and the oncogene in the WNT signaling pathway *CTNNB1*.^{37,38} Some of these mutations seem to depend on the origin of the HCC tumor, including that *CTNNB1* mutations are frequent in cases of alcohol abuse and *TP53* mutations in cases of HBV infected livers. Mutations to the activity of TERT have been shown to be the most common in HCC tumors, with around 70% of HCC cases displaying dysfunctional TERT activation, although TERT dysfunction can be derived from numerous sources including genomic amplifications, viral activity and promotor mutations.³⁹ A whole-exome genome sequencing study showed that there are 26 significantly mutated genes (SMGs) in HCC cases, with *TP53* and *CTNNB1* being the most common.⁴⁰ *TP53* in particular is of interest, considering that it is commonly seen mutated in a variety of tumor types and could play an important role in tumor initiation for HCC. Even for HCC cases without a p53 mutation, there was frequently an observed decrease in p53 activity due to alterations in activity of proteins that control p53 expression, such as MDM4. Other highly mutated genes included tumor suppressor genes *AXIN1* and *RB1* and several chromatin remodeling genes including *ARID1A*, *ARID2*, and *BAP1*. However, outside of the TERT promotor, *TP53* and *CTNNB1*, none of the analyzed genes were mutated in higher than 15% of the 363 HCC cases in this particular study. Of the 26 SMGs, the majority were found in either the WNT pathway, the PI3K pathway, chromatin remodeling, p53 signaling, or telomerase activity.⁴⁰ Among these mutated genes, it is unclear whether any particular gene is a driver of tumor development or a passenger of accompanying mutations. However, the lack of highly common mutations reinforces the

concept that HCC is a particularly heterogenous disease, which makes it difficult to treat with targeted therapies.

Due to the diversity in overall molecular signature of HCC, there has historically been great difficulty in developing and implementing molecular classifications for HCC tumors. This sort of classification, for which breast cancer is a good example, provides tremendous value in order to both offer prognoses for individual patients and to develop a treatment plan based on characteristics of a subtyped tumor⁴¹. The most significant effort to develop molecular subtypes is the Hoshida classification system, which placed HCC tumors into three distinct groupings based on multiple array-based data sets.⁴² These subtypes (S1, S2 and S3) displayed statistically significant differentiation and separated tumors into clinically relevant groups, which are summarized in Figure 1. S1 and S2 tumors displayed poorer histological differentiation, worse overall survival, and were larger in size when compared to S3 tumors. S1 tumors exhibit more vascular invasion and satellite lesions, which led to increased early recurrence after resection. Serum AFP was higher in S2 patients, which is incredibly important from a biomarker perspective. As far as molecular pathways, there are several aberrant pathways that can be considered specific to different subtypes. S1 tumors show activation of the TGF- β and WNT pathways, along with activation of E2F1 and inactivation of p53. S2 tumors display increased MYC and AKT activity, along with overexpression of IGF2. S3 tumors are mostly characterized by the maintenances of relatively normal liver function and better patient outcomes, and the most notable affected pathway is increased β -catenin expression through a *CTNNB1* mutation. Based

on differences in differentiation, invasiveness and affected pathways, it is plausible that S3 tumors arise from a different mechanism than S1 and S2 tumors. Overall, these groupings should not be considered absolute, but have shown consistency in additional studies, and match well with subclasses created using other sequencing studies.⁴⁰ The application of subtype knowledge to treatment would allow for much more specific treatment of late stage HCC patients and could be useful in making surgery decisions based on the probability of recurrence. In addition, the application of genetically based tumor information through subtyping could prove instrumental in the development of more sensitive cancer biomarkers.

Subtype	Stromal (S1)	Stemness-angiogenic (S2)	Differentiated-CTNNB1 (S3-1)	Differentiated-non-CTNNB1 (S3-2)
DNA mutations	<i>TP53</i>		<i>CTNNB1</i>	
	<i>TERT</i> promoter			
Stemness markers	<i>EPCAM</i>			
Molecular pathways	Canonical Wnt		Liver-specific Wnt	
	TGF β , interleukins, interferon	VEGF, sonic hedgehog	Xenobiotic, bile acid, fatty acid metabolism, adipogenesis	
Cellular phenotype	EMT-like, Cholangioma-like	Stem cell-like, <u>Hepatoblast-like</u>		
Histological phenotype	Less differentiated		More differentiated	
	Steatohepatic	Macro-trabecular, compact, clear cell	Micro-trabecular, pseudoglandular	
	Immune cells			
Tumor marker	AFP, GPC3			
Clinical outcome	More aggressive			

Figure 1. Hoshida HCC Subtyping System. The differentiating molecular pathways, histological phenotypes, DNA mutations, clinical outcomes, and additional markers between established HCC subtypes are outlined for the Hoshida classification system. This classification system allows for the subtyping of HCC tumors based on genetic and molecular features.⁴²

1.1.3 Treatment Options and Outcomes

With the current state of HCC treatment options, the range of outcomes for an HCC diagnosis depends heavily upon how early the cancer is detected, as there are no effective treatment options that lead to long term survival for advanced or even intermediate stage tumors. Therefore, the clinical aim is to effectively diagnose and treat small HCC lesions before they progress and metastasize. However, early stage HCC is notoriously difficult to detect with standard imaging techniques, specifically ultrasound, and therefore nodules less than 1 cm are rarely diagnosed as HCC.⁴³ The BCLC staging system, which classifies HCC cases into five stages, is the most extensively methodology of staging HCC tumors, and treatment options depend significantly on the BCLC stage of the tumor.⁸ This staging system and associated outcomes of each stage are shown in Figure 2. Patients with BCLC 0 and BCLC A have very early and early stage tumors with preserved liver functioning. These are the patients for which resection, transplantation, and ablation are viable options. At BCLC B, HCC lesions are larger (greater than 3 cm) and there are multiple nodules. Although there is typically not a deterioration in liver function or a spread beyond the liver, surgical treatments are not considered an option at this stage due to the difficulty in preventing early recurrence. Instead, chemoembolization is done to cause tumor necrosis in a relatively specific manner. At BCLC C, the tumor is at an advanced stage, meaning that there is now spread beyond the liver or cancer symptoms related to liver functioning. Treatment options at this point consist entirely of systemic therapies, which are primarily multi-kinase inhibitors. BCLC D is terminal stage HCC, and at this point patients can only

receive supportive care, with no real ability to effectively treat the cancer. These five stage guidelines inform important treatment decisions, and biomarker development efforts must aim to increase sensitivity and specificity for BCLC 0 and BCLC A tumors in order to effectively improve HCC outcomes barring the development of new treatment options for late-stage HCC.

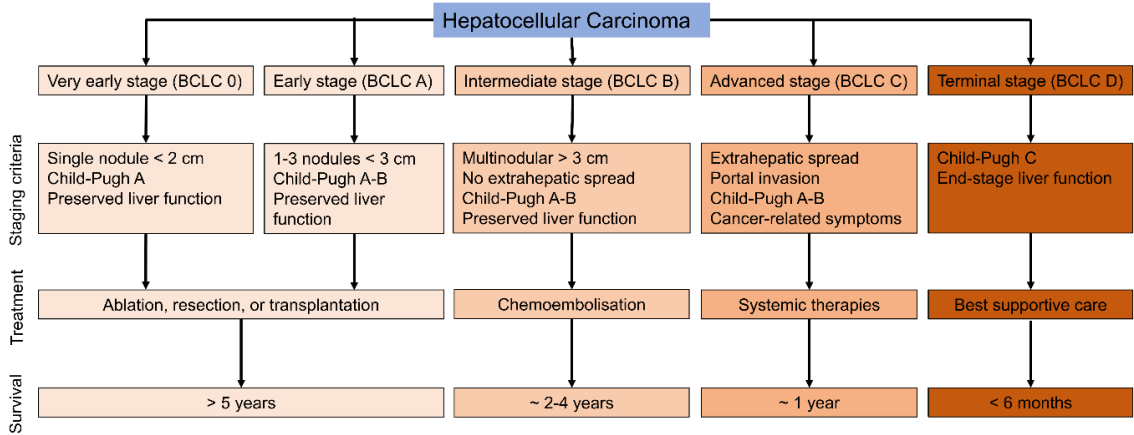


Figure 2. Hepatocellular Carcinoma Staging and Outcomes. This chart describes the Barcelona Clinic Liver Cancer staging system, with associated staging criteria, treatments, and survival outcomes.⁴⁴

Surgical treatment options for HCC have demonstrated the greatest survival benefit, but HCC must be diagnosed at stage BCLC A or BCLC 0 in order for surgery to be a common procedure. If a patient is diagnosed at an early or very early stage and is eligible for surgical options, there are a number of factors to consider when choosing between resection, ablation, and transplantation, particularly considering that HCC is a cancer typically accompanied with underlying liver conditions such as cirrhosis and hepatitis. Resection is often done in patients with particularly small HCC lesions, and is preferable in patients with minimal cirrhosis and with only a single HCC lesion. If the cirrhosis does not significantly affect liver function (Child-Pugh class A) resection can still be done, but it should not be done in patients with decompensated cirrhosis and substantially decreased liver functioning due to surgical risks. Evaluation of portal pressure and bilirubin is used to determine liver function and thereby eligibility for resection.⁴⁵ With normal bilirubin and no clinically significant portal hypertension (CSPH), patients achieve a five-year survival rate of upwards of 70% after resection. Meanwhile, patients with cirrhotic livers have five-year survivals of only 60% after resection and increased odds of both early and late recurrence.⁴⁶ Additionally, resection allows for pathologists to analyze recurrence markers, and thereby a decision for transplantation based on expected recurrence can be made before recurrence is able to advance significantly post-resection.

While resection has shown tremendous value in the correct patients, transplantation should be considered the most effective possible option, as it removes the tumor and underlying cirrhotic tissue that frequently causes late recurrence through

the de novo development of additional HCC. However, transplantation is the most invasive procedure and requires an available liver for transplant, and therefore post-surgery survival must be considered, which can eliminate some patients from eligibility. The supply of replacement livers is extremely limited, which must affect treatment decision making, particularly when considering recurrence odds. Using the Milan criteria (a single nodule ≤ 5 cm or up to three nodules ≤ 3 cm) for determining transplantation eligibility has demonstrated four-year survival of greater than 75% with recurrence rates below 15%, which is the most commonly accepted standard for transplantation.⁴⁷ Transplantation on livers with larger lesions or more than three lesions runs the increased risk of recurrence due to the possibility of extrahepatic spread prior to surgery, which is considered highly problematic due to the short supply of livers for transplantation.

Ablation is the third surgical option, in which tumor necrosis is caused by either temperature effects, typically through radiofrequency, or by injection of ethanol percutaneously. In patients with Child-Pugh A liver functioning, radiofrequency results in similar long term survival to hepatic resection in cases with less than three small lesions, and ablation has much greater cost-effectiveness.^{48,49} However, ablation cannot be performed in close proximity to surrounding organs such as the stomach and gallbladder, and ablation techniques do not allow for pathological examination for chances of recurrence. In situations in which transplantation is not viable and the liver is cirrhotic, ablation is becoming favored due to its similar survival rates and decrease in surgical invasiveness when compared to hepatic resection.

Unfortunately, the majority of HCC cases are diagnosed at either intermediate or advanced stages, which mostly eliminates surgical options due to the growth and extrahepatic spread of the cancer. TACE is the standard of care for BCLC B HCC cases, in which the aim is to take advantage of tumor-associated vascularization by delivering chemotherapy agents and blocking arterial blood supply. TACE has been shown to have survival benefits in unresectable HCC cases, and is the only form of chemoembolization for which this is true.⁵⁰ Optimal candidates for this treatment are patients that have progressed beyond surgical options, but have not yet experienced decreased liver functioning or extrahepatic spread. Among patients that adequately qualify for TACE treatment, median survival has been shown to be 48.6 months, which is the best overall survival for any treatment of HCC cases that are beyond early stage.⁵¹ TACE therapy is measured by the induction of tumor necrosis, and oftentimes needs to be repeated due to an increase in vascularization around the tumor that develops in response to treatment. Studies have shown that optimization of the delivery of the chemotherapy agent can increase survival, and one option is the use of drug eluting beads, which have been demonstrated to be highly effective in localized treatment of HCC and other cancer types.⁵¹ The use of this sort of drug delivery system minimizes off-target toxicity while standardizing treatment across patients. TACE is the only treatment shown to have survival benefits for intermediate stage HCC, and additionally there is some evidence that TACE is effective in combination with external beam radiotherapy for patients with advanced stage HCC.⁵² The BCLC C cases that have seen benefit to this combination are specifically ones with vascular invasion that do not yet have

extrahepatic spread of HCC, as this is still a therapy targeting tumor tissue locally.

Multiple studies have examined a combination of TACE therapy with systemic therapies, including sorafenib and brivanib, and there has been no survival benefit shown when compared to TACE alone for BCLC B HCC cases.^{53,54}

For BCLC C HCC cases criteria include portal invasion, impacted liver function and/or extrahepatic spread of HCC in addition to the increased tumor size seen in BCLC B cases. The only effective treatment option for cancers at this stage is found through systemic therapies, as localized treatment options such as chemoembolization have not shown to increase survival.⁸ Two orally available multikinase inhibitors, sorafenib and regorafenib, have both been shown to improve survival when compared to placebo treatment, with a median survival increase from 7.7 to 10.7 months for sorafenib and 7.8 months to 10.6 months for regorafenib.^{55,56} It was not until 2007 that sorafenib became the first systemic therapy approved by the FDA for treatment of HCC, and it is still the standard of care for advanced stage patients. Sorafenib is known to block cell proliferation and angiogenesis through interference in the RAS/RAF signaling pathway with inhibitory activity of RAF kinase and vascular endothelial growth factor receptor (VEGFR).⁵⁷ Regorafenib is structurally similar to sorafenib, and has inhibitory activity in the same kinase signaling pathways, particularly of VEGFR.⁵⁸ Several new systemic therapies have made it to clinical trials, including brivanib and sunitinib, and none have shown survival benefits over treatment with sorafenib alone.^{59,60} Regorafenib has shown added value as a secondary systemic treatment, as it has been shown to increase

overall survival from 7.8 months to 10.6 months in cases with Child-Pugh stage A liver function following previous sorafenib treatment and subsequent desensitization.⁵⁶

Based on the molecular heterogeneity of HCC, it would be expected that systemic therapies may perform differently depending on the tumor's makeup, but specific application of molecular knowledge is not currently in use in this manner in the clinical setting. There are also no systemic therapies currently in use in adjuvant settings, even though there may be value there in order to prevent recurrence, particularly following surgical treatment. There are several new drug options currently at various stages of development, including immune checkpoint inhibitors and combination therapies of antiangiogenic and immunotherapies, although none seem ready to overtake sorafenib as the first-line treatment of advanced stage HCC.⁵

In a full examination of survival outcomes for HCC, it becomes clear that the BCLC stage at the time of treatment is critical, as the systemic therapies currently available as treatments for late-stage cancers are not capable of significantly or reliably extending overall survival. With surgical treatments, there are five-year survival rates upwards of 75%, with low risk of recurrence, whereas with TACE or sorafenib five-year survival rates are much lower, with median survival under three years for TACE and under one year for sorafenib. While this clearly demonstrates the need for more effective treatments to be developed for HCC beyond surgical options, it also underscores the need to improve upon diagnostic techniques. The majority of HCC cases are diagnosed at a point at which the criteria for resection or transplantation are not met, and this fact is largely responsible for the high mortality rate of HCC.

1.1.4 Surveillance and Biomarkers

The number of high-risk patients for HCC, through any of the known risk factors, combined with the survival benefit of early detection necessitates a robust screening program for HCC in order to improve overall survival. There have been survival benefits demonstrated for HCC surveillance of patients with cirrhosis, primarily through ultrasound, yet regular screening remains inadequately utilized prior to HCC diagnosis.^{13,61} Current guidelines for screening suggest that adults with cirrhosis undergo ultrasound every six months, with or without an α -fetoprotein (AFP) serum test.³ Ultrasound alone is capable of achieving very high specificities with later stage HCC, but in order to improve overall HCC survival, more early-stage HCC cases need to be detected and treated. Ultrasound sensitivity is only 47% for early stage HCC, whereas sensitivity for early stage HCC with ultrasound and AFP in combination is 63%, which suggests that there is a benefit to using the combination.⁶² Sensitivity using ultrasound, or any imaging test, is very likely not as standardized as a serum test, and outcomes can range depending on the operator. Patient features, such as obesity or cirrhosis, can also distort ultrasound sensitivity and make diagnosis more difficult through imaging modalities.⁶³ Regardless, adherence to surveillance by ultrasound every six months is poor due to the intensive nature of requiring imaging to be done twice yearly over an extended period of time. Patients with cirrhosis need to be continually monitored, oftentimes for many years, due to an increased risk of HCC development, but this is not being done in many cases. Even among patients with diagnosed cirrhosis, over 20% are not fully adherent with surveillance by ultrasound.⁶⁴ This does not even include other

aspects of the at-risk population, such as those with metabolic disease or viral hepatitis infections. These problems with ultrasound surveillance for HCC suggest that there could be extraordinary benefit to the implementation of a serum-based biomarker with at least equal sensitivity.

There are many biomarkers that have been shown to have diagnostic ability with HCC cases, although the validation work is lacking in many cases due to difficulty in identifying early-stage HCC patients for validation studies. Serum biomarkers are a very attractive alternative to imaging surveillance due to the decreased cost and invasiveness of the procedure. AFP is the most extensively studied HCC biomarker and is the only one regularly in clinical use. However, monitoring by AFP alone has been shown to be inadequate for early and very early stage HCC monitoring, with sensitivity of 50-60% and specificity around 80% with a cutoff range around 10-20 ng/mL.^{65,66} It has been demonstrated that a significant percentage of HCC cases do not exhibit elevated AFP, and some cases with chronic liver diseases exhibit elevated AFP in the absence of HCC.⁶⁷ This means that AFP alone cannot be an effective biomarker in all HCC cases, although there could be utility in a subset of HCC cases if it was demonstrated that a subset of clearly defined cases did routinely have elevated AFP. In routine surveillance situations, it has been shown that analyzing AFP changes over time has more value than singular AFP measurements for HCC diagnosis, but this still does not fully eliminate the problems with false positives and HCC cases without elevated AFP.⁶⁸ AFP levels clearly have value when diagnosing HCC, particularly in combination with ultrasound, but AFP alone does

not have the sensitivity nor the specificity to diagnose early-stage HCC at an acceptable rate, and therefore AFP alone is not part of the HCC screening guidelines.

Beyond AFP, there are two other biomarkers with validation data beyond Phase II: AFP-L3 and des-gamma carboxyprothrombin (DCP). AFP-L3 is a specific glycoform of fucosylated AFP that has demonstrated value in detection of early stage HCC.⁶⁹ One of the main benefits of AFP-L3 is that it allows for measurements of lower serum concentrations based on a more specific assay. The primary problem with use of AFP-L3 is that AFP itself has low sensitivity, and a specific isoform of the protein would not be expected to improve on that sensitivity when used alone. This expectation holds true, as the sensitivity of AFP-L3 is decreased to around 50%, although the specificity is increased to greater than 95%.⁷⁰ This increase in specificity is particularly valuable and logically follows from AFP-L3 being a specific isoform of AFP. Combinations of AFP and AFP-L3 have shown promise, as in a recent study among early and very early HCC cases the combination of these two demonstrated sensitivity at 79% and specificity of 87% at the time of diagnosis.⁷¹ AFP-L3 showed an improved AUROC among BCLC 0 stage HCC cases in this study, which suggests that it may have value despite its low sensitivity on its own. An assay of AFP-L3 has also demonstrated value in predicting recurrence following curative treatments, which offers additional prognostic value beyond its inherent diagnostic value.⁷² Phase III validation work on AFP-L3 is still to be done, but there appears to be promise of some clinical value, although it is unlikely to be as a standalone biomarker. DCP is an abnormal prothrombin that is produced due to vitamin K uptake deficiencies in transformed hepatocytes, and it is the only other HCC

biomarker that has completed Phase II validation studies.⁷³ Sensitivity of DCP for early stage HCC is expected to range between 34-62% with specificity between 81-98%, which insinuates that it does not outperform AFP on its own.⁶⁷ DCP when combined with AFP did not improve upon AFP sensitivity and caused a decrease to AFP specificity, suggesting that a combination of these two markers would be ineffective.⁷¹ These results are in accordance with a previous study, which published that DCP is less effective than AFP with small HCC tumors and more effective than AFP with large HCC tumors.⁷⁴ Due to the pressing need to diagnose and treat early stage HCC cancers, this quality detracts from DCP's clinical value and makes it likely that DCP would only provide diagnostic value in combination with another biomarker, although it has already been shown not to enhance specificity or sensitivity of AFP.

There are a number of biomarkers that have shown at least somewhat promising results in early clinical validation for diagnosis of early-stage HCC but have not yet progressed into Phase III longitudinal studies. Osteopontin (OPN) is an integrin-binding phosphoprotein that has been shown to be expressed in a multitude of cancer types including pancreatic, colon, and HCC. OPN has shown promise as a biomarker for several of these cancers, with multiple studies comparing its diagnostic value to AFP.⁷⁵ In a meta-analysis of early-stage HCC serum, OPN showed a pooled sensitivity of 49% and pooled specificity of 71%. Combining with AFP led to an increase to 73% in sensitivity with similar specificity of 68%.⁷⁶ OPN has similar diagnostic value to AFP in both early-stage and all HCC cases, and there may be some value to a combination of OPN and AFP. Midkine (MDK) is a heparin binding growth factor that has shown diagnostic potential

for early-stage HCC cases. MDK has been shown to have significant diagnostic value among BCLC 0/A cases, with a sensitivity as high as 87%, and among HCC cases with normal AFP levels.⁷⁷ This would suggest that MDK combined with AFP would have significant diagnostic value, which has been shown to be true, although more longitudinal studies are needed to confirm the sensitivity across HCC cases.⁷⁸ GP-73 is a golgi-specific transmembrane glycoprotein that has been shown to be elevated and have altered glycosylation patterns in HCC cases.⁷⁹ The sensitivity of GP-73 for BCLC 0/A HCC cases has been shown to be 62%, with a specificity of 88%, which makes it competitive with AFP.⁸⁰ However, it has proven difficult to develop a GP-73 ELISA-based assay, which has limited its application to clinical use or further validation. There are a variety of other biomarkers associated with HCC, including dickkopf-1 (DKK1), glypican-3 (GPC-3), alpha-1 fucosidase, and squamous cell carcinoma antigen, but all of these have problems related to performance against early stage HCC, low specificity, a lack of value in a combination with AFP, and/or outperformance by AFP alone.^{67,81-84}

In addition to the diagnostic value demonstrated in the discussed biomarkers, including the complementary use of two or more of them, there have been a multitude of algorithms developed with clinical value. The use of an algorithm, particularly in the case of a disease with many molecular pathways that are potentially dysfunctional, allows for enhanced specificity through the analysis of a greater portion of the dysregulated tumor biology. Algorithms are able to include multiple biomarkers in addition to clinical and etiological factors, which in theory enables significantly more precision. Potentially the most validated diagnostic panel for HCC is the GALAD score,

which includes gender, age, AFP, AFP-L3, and DCP.⁸⁵ This model has been developed from large data sets of patients in the UK, Germany and Japan, and it has shown ability to distinguish between HCC and other chronic liver diseases. For early stage HCC cases, GALAD has demonstrated sensitivity ranging from 80-83% and specificity from 82-89%, which outperforms any of the serological biomarkers on their own.⁸⁶ Another potentially useful algorithm that has undergone validation work for HCC diagnosis is the Doylestown algorithm, which attempts to diagnose HCC based on log AFP, gender, age, alkaline phosphatase, and alanine aminotransferase.⁸⁷ In the initial discovery of the algorithm, at a fixed specificity of 95%, detection of HCC was improved by 2-20% when compared to AFP alone. The Doylestown algorithm has shown sensitivity greater than 50% up to a full year before clinical HCC diagnosis, which outperforms AFP alone and thereby demonstrates potential clinical value for early stage HCC.⁸⁸ Furthermore, the addition of fucosylated kininogen to the algorithm demonstrated a sensitivity of 89% for early stage HCC with normal AFP levels.⁸⁹ This algorithm also needs further validation from longitudinal studies with early stage HCC, but it demonstrates potential diagnostic value for early stage HCC regardless of etiology.

1.2 N-linked Glycosylation

1.2.1 Broad Overview of Glycosylation

Glycosylation refers to the addition of monosaccharide or oligosaccharide chains, commonly referred to as glycans, to protein or lipid molecules in order to modulate both size and function. There are nine common monosaccharides that form glycan structures in vertebrates, and these can be grouped by their molecular weight. Hexose (Hex) monosaccharides consist of glucose (Glu), galactose (Gal), and mannose (Man) residues, which have an $m/z = 162.0528$. N-acetylhexosamine (HexNAc) residues include an acetyl addition to a hexose, and include N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc) monosaccharides with an $m/z = 203.0794$. Additional monosaccharides include fucose (dHex, $m/z = 146.0579$), N-acetylneuraminic acid (NeuAc, sialic acid, $m/z = 291.0954$), xylose (xyl, $m/z = 132.0423$), and glucuronic acid (GlcA, $m/z = 176.0321$). Figure 3 displays the representing depictions that will be used in place of monosaccharide chemical structures for simplicity throughout this work. Some of these monosaccharides are only expressed in certain species or glycoforms. For example, xylose residues can be expressed in complex glycans in plants, but are only found on proteoglycans in humans.⁹⁰ Although the number of monosaccharides are limited, there is still significant glycan structural complexity due to a lack of a rigid template to encode glycan synthesis. Substrate availability, enzyme localization, pH, and availability of cellular nutrients all can alter the expressed glycoforms, which allows for dynamic glycosylation with changing biological conditions.⁹¹

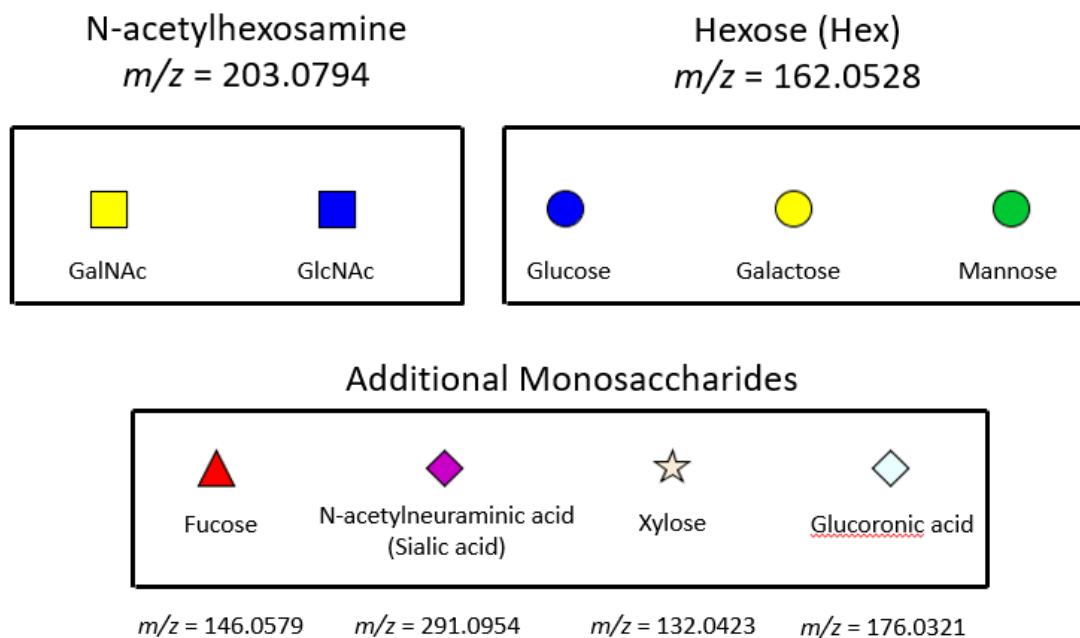


Figure 3. Monosaccharide Structural Representations. Depicted are the representative images of each monosaccharide that will be used in place of chemical structures for simplicity.

As a post-translational modification, glycosylation adds valuable complexity to the expression of protein and lipid structures, allowing for increased control over activity dynamically. Although common structural motifs exist, there are numerous potential glycan structures that can be variably added to glycan attachment sites. Glycan branching, particularly on N-glycans, presents even more complexity as differing structures can be assembled with the same monosaccharide additions. It has been estimated that more than 3000 N-glycan structures are possible within the human N-glycan serum proteome.⁹² While biosynthetic pathways force glycans to be assembled into easily definable structures, there is still enormous variability based on potential size, linkage differences, branching, fucosylation, and sialylation, along with additional modifications such as sulfation. Without this structural complexity, it would be impossible for glycans to be the multi-faceted participants in cellular functioning that they need to be for healthy cell growth and signaling.

In humans, the primary types of glycosylation include N-linked glycosylation, O-linked glycosylation, glycosphingolipids, and proteoglycans.⁹³ N-glycosylation refers to the attachment of an N-acetylglucosamine (GlcNAc) to an asparagine residue of a glycoprotein at an amino acid sequence of Asn-X-Ser/Thr where X is an amino acid but proline.⁹⁴ N-glycans have a GlcNAc₂Man₃ core structure, with a variety of possible monosaccharide additions to modulate structure and function. O-glycosylation involves the similar addition of GlcNAc (or GalNAc) to glycoproteins, but attached to serine or threonine residues instead. O-glycosylation is highly diverse structurally and can be classified based on the initial sugar attachment, such as mucin-type O-glycans for

GalNAc attachments.⁹⁵ Glycosphingolipids are formed from the addition of a glycan to the ceramide of a sphingolipid, and they have important structural roles in the formation of lipid rafts.⁹⁶ Proteoglycans are extracellular glycoproteins that have elongated sugar repeats termed glycosaminoglycans attached via O-linked glycan motifs.⁹⁷ While complex N-linked and O-linked glycans typically contain 5-12 sugar monosaccharides, glycosaminoglycans can contain up to 80 sugars, making them much larger structures.⁹³ These large structures are essential for the formation of the glycocalyx, which is essential for extracellular functioning.

Glycans have a wide range of necessary cellular functions that are critical for signaling, growth, and survival. Each structural classification of glycans has distinct cellular functioning, although there are many roles that require coordination across glycan subclasses. The functions of glycans can be approximately classified into four groupings: structural roles, intrinsic recognition, extrinsic recognition, and host mimicry.⁹⁸ Structural roles of glycans are numerous and include functions such as the creation of a physical barrier, glycoprotein folding, protection from protease degradation, and the modulation of receptor signaling at the membrane.^{99–103} Intrinsic recognition describes intraspecies recognition of glycans within an organism. Examples of intrinsic roles of glycans include glycoprotein trafficking, intercellular signaling, and intercellular adhesion.^{104–106} Meanwhile, extrinsic recognition describes interspecies glycan recognition between hosts and pathogens/symbionts. Examples of extrinsic roles of glycans include bacterial adhesins, bacterial toxins, and pathogen

recognition.^{101,107,108} Host mimicry describes the appropriation of host glycans for the functioning of microorganisms.^{109,110}

1.2.2 Synthesis and Structure of N-Glycans

N-glycosylation specifically refers to the attachment of a GlcNAc to the nitrogen of an asparagine residue found within the amino acid sequence Asn-X-Ser/Thr where X denotes any amino acid excluding proline. Following the covalent bond attachment of a GlcNAc, various monosaccharides including galactose, mannose, fucose, sialic acids, and additional GlcNAcs can be added/removed to form N-glycan structures. After biosynthesis, the final structure forms either a complex, hybrid, or high mannose N-glycan, examples of which are illustrated in Figure 4. There are many possible structures within these categories, particularly for complex structures, depending on the galactosylation, number of antennae, decoration with fucose/sialic acid, and additional modifications such as sulfation. N-glycans are necessary for both intracellular and extracellular functions, with new mechanisms of action still being discovered regularly.¹¹¹

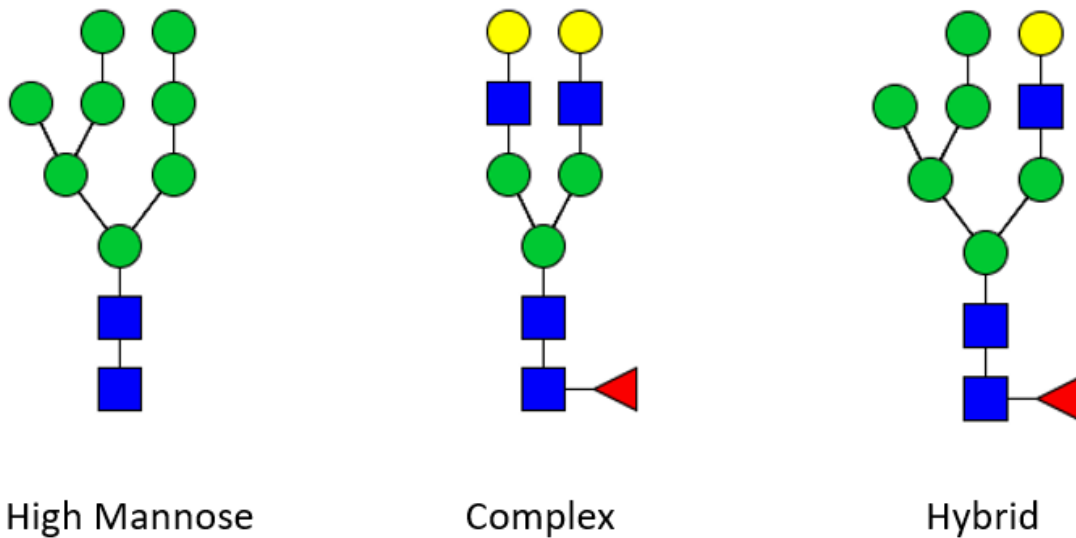


Figure 4. N-Glycan Structural Classifications. Illustrated are the three categorical classifications of N-glycan structures. High mannose N-glycans have only mannose monosaccharides beyond the $\text{GlcNAc}_2\text{Man}_3$ core structure. Complex N-glycans have two or more antennae with a GlcNAc attached to the $\text{GlcNAc}_2\text{Man}_3$ core structure. Hybrid N-glycans have only mannose monosaccharides attached to half the $\text{GlcNAc}_2\text{Man}_3$ core structure, and one or more antennae with a GlcNAc attached to the other half of the $\text{GlcNAc}_2\text{Man}_3$ core structure.

N-glycan biosynthesis in eukaryotic cells occurs in two phases, located in the endoplasmic reticulum (ER) and in the Golgi, where oligosaccharides are transferred to Asn and processed by glycosidase and glycosyltransferase enzymes.^{93,94,112,113} In the endoplasmic reticulum, an N-glycan precursor is synthesized and attached to dolichol phosphate (Dol-P). Dolichol is a polyisoprenoid lipid necessary for oligosaccharide transfer to nascent glycoproteins. The mature precursor is $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$, and once it is synthesized through the use of glycosyltransferases and UDP-GlcNAc, GDP-Man, and UDP-Glc substrates it can be transferred to Asn by oligosaccharyltransferase (OST). This transfer happens co-translationally as the protein is being translocated into the ER. Further processing in the ER includes the glucosidase catalyzed removal of glucose monosaccharides that serves as a control over correct glycoprotein folding, along with mannosidase catalyzed removal of a mannose.¹¹⁴

The glycoprotein is then moved to the Golgi, where further processing results in complex and hybrid N-glycans. In the *cis*-Golgi, the $\text{Man}_8\text{GlcNAc}_2$ oligosaccharide is trimmed by mannosidases to a $\text{Man}_5\text{GlcNAc}_2$ structure, which is the key intermediate for complex/hybrid type N-glycans. If a glycan is to retain high mannose structure, it will leave the Golgi at some point during mannosidase trimming. In the *medial*-Golgi, glycosyltransferases (referred to as MGAT enzymes) can add GlcNAc monosaccharides to form terminal antennae onto the base structure, which results in complex or hybrid N-glycans. MGAT1 and MGAT2 are responsible for GlcNAc additions to the α 1-3 and α 1-6 mannose residues to form the base structure for complex biantennary N-glycans. Additional glycan branching can be initiated through MGAT3 (bisecting N-glycans) and

MGAT4/5 (triantennary/tetra-antennary N-glycans). Each antenna can then be extended through the action of galactoses along with potential sialic acid and/or fucose decoration. Fucose can also be added to the core GlcNac through a α 1-6 linkage. The decoration of complex N-glycans depends on the availability of substrate, the expression of needed glycosyltransferases, and the biological demands of the cell.

1.2.3 Dysregulated N-Glycosylation in Cancer

The link between altered glycosylation and cancer initiation and progression has been long established, which includes the fact that many of the clinically utilized cancer biomarkers are glycoproteins. However, the mechanistic link between aberrant glycosylation and tumor progression is still not very well understood, and it appears as though molecular and genetic features of tumors play an important role in specifically how glycosylation is dysregulated.¹¹⁵ There have been a number of studies regarding dysregulated glycosylation in a variety of tumor types, with varied glycosylation alterations observed in both a protein-specific and cell-specific manner.¹¹⁶ Aberrant glycosylation in cancer can emerge based on a few driving factors, mainly related to either the expression/localization of glycosyltransferase enzymes and/or the availability of needed sugar nucleotide donors.^{117,118} Although there is some variance between differing cancer types, the most commonly occurring cancer-related glycan changes are related to sialylation, core fucosylation, O-glycan truncation, and O- and N-glycan branching.^{119–122} Aberrant cancer glycosylation signals that glycosylation is critical for

healthy cell functioning, and that it plays important roles in cell proliferation and cell growth.

There have been a number of glycan alterations identified in different cancer types, many of which are either already directly related to an established biomarker, such as the fucosylated glycoform of AFP AFP-L3, or showing promise in application to biomarker development.⁷⁰ Glycan alterations have been identified in the majority of studied cancer types, emphasizing their relevance in cancer progression. One of the most commonly observed glycan alterations in cancer is related to sialylation, which are terminal sugars on glycoconjugates that play valuable roles in cell to cell signaling, cell adhesion, and cellular recognition.^{94,120} Sialylated Lewis epitopes are well-known to be related to cancer progression, and have been shown to be related to metastasis through cell adhesion effects, such as the increase in sLe^x in breast cancer metastases.^{123,124} This trend is present in additional tumor types, with sLe^x being correlated with poor prognosis in colon cancer and with metastasis of ovarian and liver cancers.^{120,125} Apart from Lewis antigens, an increase in overall sialylation has been linked directly to cancer, particularly α 2-3 and α 2-6 linkages.¹²⁶ Sialylated glycans have been shown to be increased in a variety of tumor types, including increased expression on N-glycans of pancreatic cancer and increased expression of sialyltransferases for both N- and O-linked glycans in breast cancer.^{127,128} α 2-6 sialylation of β 1 integrin has been directly linked to metastasis of ovarian cancer cells.¹²⁹ Overall, it is clearly established that sialylated glycoconjugates play an important part in cancer metastasis in a number of cancer types, including breast, liver, ovarian, pancreatic and colon.

In addition to sialylation, fucosylation is the other major glycan alteration relevant in cancer related to the expression of particular sugar subunits on glycoconjugates. Fucosylated glycans can be separated into core fucosylated and terminally fucosylated structures, with fucosyltransferases (FUT) responsible for the fucosylated glycan biosynthesis. There are 11 FUTs for N- and O-linked fucosylated glycan synthesis, although FUT8 is solely responsible for core fucosylation (α 1-6 linkages).¹¹⁹ Core fucosylation has been shown to be increased in tumor tissue in breast, melanoma, liver, ovarian, colon, pancreatic, and lung cancers.^{119,130-134} In breast cancer, core fucosylation of epidermal growth factor receptor has been correlated with tumor cell growth.¹³⁵ Core fucosylation is associated with liver cancer development, and it is increased directly on a number of serum glycoproteins.¹³⁶⁻¹³⁸ Overexpression of FUT8 has been directly observed in breast and liver cancers.^{139,140} Additionally, Le^x antigens are differentially expressed in breast and colon cancers.¹²⁰ Overall, fucosylation is important for tumor cell growth and invasion, and aberrant fucosylation is clearly a distinguishable hallmark of cancer when compared to non-cancerous tissue.

A primary driver of aberrant glycosylation in cancer is the expression and activity of glycosyltransferases responsible for biosynthesizing particular glycan structures, particularly sialylated, core fucosylated, and branched glycans.¹⁴¹ Core fucosylated glycans have been well established to be relevant in cancer progression, and the only enzyme responsible for their biosynthesis is FUT8. In liver cancer, it has been shown that there is an amplification of genes related to the *de novo* synthesis pathway of GDP-fucose, which is a key substrate for FUT8 activity.¹³⁷ Knockout of FUT8 in rodent models

has inhibited HCC tumorigenesis, although mice displayed growth retardation and lung emphysema.¹⁴² In lung cancer, it has been shown there is upregulation of FUT8 in non-small cell lung cancer (NSCLC) which correlates with metastasis and disease recurrence.¹⁴³ There are multiple sialyltransferases that add sialic acids to glycoconjugates in a linkage dependent manner, the expression and/or activity of which have shown to be dysregulated in many cancers. Both α -2,3-sialyltransferase 1 (ST3Gall) and α -2,6-sialyltransferase 1 (ST6Gall) have been shown to be upregulated in several cancers, including colon and breast.^{144,145} For these reasons, both sialyltransferases and fucosyltransferases have made for intriguing therapeutic targets, although there is still significant progress needed before such inhibitors have widespread clinical value.¹⁴⁶ In addition to sialyl- and fucosyltransferases, glucosaminyltransferases that catalyze the biosynthesis of bisecting and branched N-glycans are commonly dysregulated in cancer. Rodent knockout models of golgi β 1,6N-acetylglucosaminyltransferase V (MGAT5), which is responsible for tetra-antennary N-glycan biosynthesis, have shown that MGAT5 suppression leads to decreased tumor growth and metastasis.¹⁴⁷ Branched glycans biosynthesized by MGAT5 have been well-established to result in decreased cell-cell adhesion and increased tumoral invasion.¹⁴⁸ The expression of these branched glycan structures has been directly related to cancer metastasis and survival, through a variety of mechanisms.¹⁴⁹

In cancer cells, abnormal glycosylation plays a critical role in proliferation and survival through promoting metastasis/invasion, promoting angiogenesis, promoting tumor proliferation, and avoiding anti-tumor immune responses.¹⁵⁰ It has been

thoroughly demonstrated that glycosylation is a primary regulator of growth and death factor receptor signaling, and core fucosylation of these receptors can promote cancer growth and proliferation.¹⁵¹ The mechanisms through which glycosylation affects the activity of these receptors includes through galectin binding, which can decrease their turnover on the plasma membrane and thereby promote growth factor receptor activity.¹⁵² The formation of galectin-binding lattices that promote receptor activity is often driven by MGAT5 biosynthesis of β 1-6 branched N-glycans. In contrast, the addition of a bisecting GlcNAc to N-glycans increased the endocytosis of epidermal growth factor receptor (EGFR) and thereby decreased proliferative signaling.¹⁵³ Another mechanism by which glycosylation affects cancer progression is through the overexpression of sialic acids and its effect on cell-cell and cell-matrix adhesion. Through the expression of these negatively charged sialic acids, cell-cell adhesion is decreased and tumor cell migration is increased in breast cancer cells.^{154,155} Glycosylation is also relevant in cancer cell metabolism, as increased glucose levels lead to increased availability of UDP-GlcNAc, which is a critical substrate for both O- and N-glycosylation. Knockdown of O-GlcNAc transferase *in vitro* has led to inhibited tumor growth and metastasis.¹⁵⁶ Another important mechanism to consider in regard to tumor cell glycosylation is immune response and the escape of immune surveillance by cancer cells. There are various lectins that are able to bind glycans and thereby regulate immune processes, and the expression of abnormal complex glycans may allow cancer cells to evade such immune responses.^{157,158} Avoidance of immune effector cells is critical for tumor progression, and it has been clearly demonstrated that this avoidance

is modulated by aberrant glycosylation. Overall, these are just a few examples of the many ways in which aberrant glycosylation can mechanistically lead to increased tumor cell proliferation and metastasis, and more research is needed to further elucidate additional mechanisms.

1.2.4 N-Glycosylation in Cancer Diagnostics

There have been many cancer biomarkers developed in an attempt to reduce cancer mortality through early diagnosis, and the majority of these markers are serum glycoproteins. Interestingly, many of these markers do not consider the glycan content on the protein, but nevertheless this development points to the importance of glycoproteins within cancer development and diagnosis. The incorporation of glycan-based cancer biomarkers into cancer diagnostics is gaining in popularity in research settings, although clinical implementation has been slow. In order to clinically implement cancer biomarkers, diagnostic sensitivity must be demonstratively improved and clinical implementation must be minimally invasive. This is a major benefit for glycan-based biomarkers, as serum glycoprotein biomarkers are much simpler to clinically implement than any sort of biopsy-dependent biomarker.

Perhaps the most widely implemented of the glycoprotein cancer biomarkers is prostate-specific antigen (PSA) for prostate cancer.¹⁵⁹ It has long been established that the recommended methodology for surveillance for prostate cancer is a PSA blood test with or without a digital rectal exam (DRE), which then can lead to a prostate biopsy to confirm the cancer. However, questions have been raised regarding the efficacy of

serum PSA in distinguishing between prostate cancer and noncancer prostatic disease, which leads to many unnecessary prostate biopsies.^{160,161} This lack of specificity is in large part why serum PSA screening is typically done in conjunction with a DRE. However, recent studies have shown that differentially glycosylated PSA may be able to serve as a more specific marker to differentiate benign prostatic disease from prostate cancer.^{162,163} This is consistent with previous findings demonstrating increased core fucosylation and α 2-3 sialylation in the whole serum N-glycome of prostate cancer compared to benign prostate hyperplasia.¹⁶⁴ Thus far, core fucosylated PSA was successful in differentiating prostate cancer from benign prostate hyperplasia, but not in differentiating aggressive prostate cancer from non-aggressive prostate cancer (determined by Gleason score).¹⁶⁵ PSA is a successful, clinically used biomarker for prostate cancer, although low specificity limits its overall effectiveness, which is a problem that may be addressed with further research by utilizing specific glycoforms of PSA.

Another cancer with a very commonly used glycoprotein serum biomarker is AFP for the diagnosis of HCC. Currently, as previously discussed, clinical guidelines rely on abdominal ultrasound with optional AFP testing which leads to many tumors being diagnosed at late stages.¹⁶⁶ HCC has a very defined risk group, as the vast majority of tumors develop in patients with cirrhotic livers, which makes the group in need of cancer surveillance very clear. A combination of AFP and ultrasound has been shown to improve sensitivity for early stage HCC over ultrasound alone.^{62,167} Unfortunately, up to 40-50% of HCC tumors do not have elevated AFP levels, making the sensitivity for early

stage HCC of AFP inadequate to improve clinical outcomes substantially on its own, even if there is an added benefit of combined ultrasound/AFP screening.^{66,67} However, AFP-L3, a fucosylated glycoform of AFP, has been shown to potentially improve detection of early stage HCC.¹⁶⁸ The major limitation is that an isoform of a low sensitivity biomarker is unlikely to dramatically improve sensitivity, although the increased specificity makes utilization of AFP-L3 appealing for algorithm-based diagnostic approaches. Both AFP and PSA have glycoforms that enhance its diagnostic performance, which highlights the capability of glycan-based biomarkers for the improved diagnosis of cancer, particularly at early stages.

AFP and PSA are the most frequently utilized cancer biomarkers clinically, although there are a number of other glycoprotein standalone biomarkers that have been identified in research and used to some degree clinically. CA125 has been established for use in both diagnosing and monitoring the progression of ovarian cancer.¹⁶⁹ CA125 is not sensitive enough to reliably detect early stage ovarian cancer, as it is elevated in only 50% of stage 1 cancers, which are the ones that are most treatable with surgical cure.¹⁷⁰ However, it has clinical value in predicting recurrent ovarian cancers, as CA125 levels rise at least 3 months before recurrence and two months before diagnosis through imaging.¹⁷¹ The sensitivity for recurrence is 62-94% and the specificity is 91-100%.^{170,172} While it may not have early stage diagnostic value, CA125 is clearly clinically useful for monitoring ovarian cancer. CA15-3 has been associated with a variety of different types of cancer cells, but it is most predominantly used in the post-operative monitoring of breast cancer.¹⁷³⁻¹⁷⁵ CA 15-3 monitoring has shown clinical

utility in monitoring the potential for recurrent breast cancer and in the response to ongoing treatment.¹⁷⁶ In recent years, MUC1 has gathered attention as an attractive potential target for developing immunotherapies for breast cancer.¹⁷⁷ CA 19-9 is another well-known glycosylated tumor marker, which is primarily used for the monitoring of pancreatic cancer, although it has also been examined as a marker of colorectal, gastric, and liver cancers.¹⁷⁸ CA19-9 is approved by the FDA for diagnosis of pancreatic cancer, although the utility in diagnosing early-stage pancreatic tumors is contested. It has been shown that CA19-9 can have relatively high sensitivity for pancreatic cancer, with one study showing a sensitivity of 68% a full year before diagnosis.¹⁷⁹ However,, CA19-9 can frequently be increased in patients with non-malignant conditions, requiring high cut-off values for adequate specificity and limiting its clinical utility.¹⁸⁰ This need for high cut-ff values is a severe limitation for its clinical utility. CA19-9 can also be used to monitor potential recurrence following surgical treatment of pancreatic cancer.¹⁸¹ Interestingly, recent work has shown that related glycans to the CA19-9 antigen have sensitivity for distinct subsets of pancreatic cancers, showing the value of both utilizing glycomics within biomarker development and targeting genetically/molecularly similar subgroupings of tumors for biomarker development.¹⁸²

In addition to these biomarkers with established clinical utility in either cancer diagnosis/surveillance or post-treatment cancer monitoring, there is a number of serum glycoproteins that have been examined in research settings to have cancer biomarker potential. α -1-Antitrypsin (A1AT) is a serum glycoprotein primarily produced by liver

hepatocytes that has been associated with multiple cancer types, including liver and lung. There has been evidence that differentially glycosylated glycoforms of A1AT in serum can be utilized to distinguish between non-small-cell lung cancer from benign pulmonary diseases and to distinguish lung adenocarcinomas from benign diseases.¹⁸³ Core fucosylation of A1AT has been directly linked to the progression of liver cirrhosis to hepatocellular carcinoma.¹³⁶ Fucosylated kininogen is another serum glycoprotein that has been shown to have increased expression in hepatocellular carcinoma compared to cirrhosis, and increases diagnostic sensitivity particularly when used in combination with AFP.¹⁸⁴ A fucosylated glycoform of haptoglobin has been shown to be increased in a number of cancer types, including pancreatic, breast, and liver.¹³⁸ Additionally, increased fucosylated haptoglobin correlates with a poorer survival and increased metastasis in these cancer types.¹¹⁹ The cancer-associated increased fucosylation, particularly core fucosylation, of a number of serum glycoproteins, is validated by an observed amplification of genes involved in the *de novo* GDP-fucose synthesis pathway for liver cancer samples.¹³⁷

Due to the inadequate sensitivity of a number of standalone cancer biomarkers, it is becoming much more common to develop cancer biomarker algorithms, which can incorporate multiple biomarkers along with relevant clinical and/or demographic information. These algorithms tend to provide increased sensitivity based on biomarkers that are complementary, although they can be more difficult to implement clinically due to the increased complexity and need for additional information. These sort of biomarker algorithms have been developed for a number of cancer types, including

liver, ovarian, and pancreatic. For liver cancer the most well-known biomarker algorithm is the GALAD score, which combines age, gender, AFP-L3, AFP, and DCP to improve upon diagnostic sensitivity and specificity of AFP alone.⁸⁵ For early stage hepatocellular carcinoma, GALAD has been demonstrated to have sensitivity from 80-83% and 82-89%, which is an improvement on any of the known standalone serum biomarkers.⁸⁶ The GALAD score shows great promise in improving clinical outcomes through improved detection, although more validating work is required before it can be widely implemented into the clinical setting. Similarly to the GALAD score, the Doylestown algorithm aims to distinguish HCC from chronic liver disease utilizing a number of biomarker/demographic factors, including AFP, gender, age, alkaline phosphatase (ALK), alanine aminotransferase (ALT), and recently added fucosylated kininogen.⁸⁸ The Doylestown algorithm has shown comparable early-stage HCC detection to the GALAD, with true-positive rates (at 10% screening false positive level) of 63.2% and 57.9% respectively, both of which are improvements on the sensitivity of AFP alone for the same cohort.¹⁸⁵ For ovarian cancer, the risk of ovarian malignancy algorithm (ROMA) has been developed to predict the risk of ovarian cancer in patients with benign ovarian masses.¹⁸⁶ This algorithm utilizes CA125 and HE4, along with menopausal status, in order to stratify patients into high and low risk groups, and a sensitivity of 92.3% was demonstrated for a postmenopausal group.¹⁸⁷ Compared to using CA125 alone, this algorithm doubled the number of tumors being accurately diagnosed in a large scale study in the United Kingdom.¹⁸⁸ For pancreatic cancer, it is critical to correctly identify cancerous and noncancerous cysts in order to avoid overdiagnosis of cancer. Utilizing a

biomarker risk score that includes VEGF, glucose, and CEA, 91% of cysts were correctly identified preoperatively, which is a significant improvement over current clinical outcomes.¹⁸⁹ Algorithms such as these show substantial promise to improve diagnostic sensitivity/specificity over standalone biomarkers or current clinical diagnostic techniques, and they should be seriously considered for biomarker development moving forwards.

1.2.5 Glycomics Methods of N-Glycan Analysis

1.2.5.1 Biosynthetic Inhibitors of N-Glycans

Historically, the most common methodology to analyze glycan function was through the utilization of glycan biosynthetic inhibitors, which are able to control the expression of glycans in cell culture. Tunicamycin, a complete N-glycosylation inhibitor, is the most well-known and is commonly used as a tool to analyze the effect of N-glycans on cellular functioning.^{190,191} Tunicamycin functions through the inhibition of GlcNAc-1-phosphotransferase, which catalyzes the essential addition of a GlcNAc monosaccharide onto dolichol for the biosynthesis of N-glycans. Tunicamycin was vital in assessing the critical role that the N-glycan biosynthetic pathway plays in correct glycoprotein folding. Various other small molecule inhibitors that block glucosidases and mannosidases essential to N-glycan biosynthesis have also been discovered including castanospermine, kifunensine, and swainsonine.¹⁹²⁻¹⁹⁴ This group of inhibitors blocks the progression along the biosynthetic pathway to complex N-glycans, resulting in primarily high mannose type N-glycans. Additionally, there are inhibitors to block the

addition of capping monosaccharides, such as 2-deoxy-fluorofucose (2FF) to inhibit fucosylation and a number of CMP-sialic acid analogues to inhibit sialylation.^{146,195–197} In summary, there are small molecule inhibitors for nearly every critical step in N-glycan biosynthesis, which provides experimental control over N-glycan expression. While the cross-toxicity of these inhibitors has rendered their clinical value to be extremely limited, there is a tremendous amount of experimental value in utilizing these inhibitors to assess N-glycan functioning in various cellular states. It is difficult to gain a nuanced understanding through the use of complete inhibitors such as tunicamycin, but these inhibitors have nonetheless provided critical knowledge about N-glycan structure and functioning.

1.2.5.2 Lectin-based Analysis of N-Glycans

The existence of glycan-recognizing probes (GRPs), including lectins, was first established in conjunction with the analysis of human blood group antigens.¹⁹⁸ At this point in time, hundreds of GRPs have been identified, with various affinities for different glycan monosaccharides/structural motifs. Lectin affinity chromatography has long been established as an effective way to analyze complex glycopeptide mixtures, although they are highly dependent on the availability and specificity of GRPs.¹⁹⁹ Lectins are commonly used as an enrichment strategy, using affinity chromatography to isolate glycopeptides of interest, then using peptide-N-glycosidase F (PNGase F, an enzyme that cleaves N-glycans from asparagine) to deglycosylate captured glycopeptides for glycan analysis.²⁰⁰ The majority of lectins are originally derived from plants, and offer affinity in

the low micromolar range for complex glycans. With such affinities, the utility of lectins is to identify broad glycan structural motifs within a sample set, and if more precise structural features are needed it is best to couple lectin enrichment with an orthogonal approach such as mass spectrometry.

There are a number of lectins commonly used in N-glycan analysis, all with varying affinities for various N-glycan structural elements.²⁰¹ Depending on the needed application, there are a number of lectins to choose from with specific binding affinities to differing N-glycan structural features.²⁰² Some lectins preferentially bind mannose, including *Concanavalin-A* (Con A), which binds terminal mannose and biantennary structures, and *Arum maculatum* (AMA), which binds N-acetyllactosamine and high mannose type glycans.^{203,204} Con A is perhaps the most widely used lectin, as it has broad specificity for glycoproteins containing N-glycans. There are also lectins that preferentially bind complex N-glycans, including *Colchicum autumnale* (CA) to bind biantennary structures, *Phaseolus Vulgaris* (PHA-E) to bind bisecting structures, and *Phaseolus Vulgaris* (PHA-L) to bind β 1-6 branched structures.^{205,206} Lectins to bind capping monosaccharides are also commonly used, including *Aleuria aurantia lectin* (AAL) to bind fucose and *Sambucus nigra Agglutinin* (SNA-I) to bind α 2-6 sialic acid.^{207,208} Additionally, it is possible to engineer lectins with enhanced binding properties to specific structural features, such as recombinant AAL lectins for core fucosylated glycans.^{209,210}

One of the most popular experimental utilizations of lectins is through a lectin microarray, in which immobilized lectins capture glycans for analysis in a high-

throughput manner.²¹¹ Lectin microarrays have numerous advantages for glycomics analyses, including an ease of sample processing through fluorescent tagging, identification of glycan linkages through lectin specificity, and minimal need for sample volume. Due to the wide availability of diverse GRPs, the applicability of lectin microarrays is enormous in the field of glycoproteomics. Once again, this application is limited by the specificity of the GRP in use, and it can rarely isolate single glycan structures with precision. However, it is a valuable tool for researchers aiming to analyze glycan structural features of a glycoprotein sample set. Additionally, as GRPs with more specific binding affinities are engineered, the experimental value of lectin microarrays grows. There are now many commercially available lectin microarrays, although they are also commonly designed in-house depending on experimental needs and conditions.

1.2.5.3 High Performance Liquid Chromatography of N-Glycans

In order to gain specific structural information on N-glycans, a common approach is to use liquid chromatography (LC) separation to isolate and identify specific N-glycan structures. In the early days of glycomics, utilizing LC for glycan identification was difficult, because glycan mixtures are often very complex with multiple glycans of similar sizes, shapes, and charges. However, this issue has been largely mitigated by technological advances, particularly in ultra-high performance liquid chromatography (UPLC), making LC a viable experimental route of glycan identification and analysis.²¹¹ To perform UPLC analysis, N-glycans are released from associated glycoproteins through

enzymatic cleavage, traditionally PNGase F. From this point, glycans are labeled with a fluorescent tag, such as 2-aminobenzamide (2-AB), for detection. Retention time can often vary based on column conditions and flow rate, although this can be overcome by standardizing to glucose units (GU) through the use of a glucose homopolymer ladder. Thereby each N-glycan of interest will elute at a specific GU based on the eluent and accurate glycan identification can be done. However, this is highly reliant on the development of accurate databases of GU values for glycans of interest.

Depending on the LC conditions and the complexity of the sample mixture, there is often not an adequate separation of N-glycans, leading to identification and quantification issues. In order to overcome this, glycosidases are commonly used in order to sequentially digest N-glycans, which alters their GU in a digestion-dependent manner.²¹² Common glycosidases used for this purpose include neuraminidases (otherwise termed sialidases), fucosidases, galactosidases, and mannosidases. Through such methodologies, UPLC analysis is viable in correctly separating and identifying many complex glycans, oftentimes including variable glycan linkages. It is feasible to separate β 1-4 branching from β 1-6 branching, as well as α 1-6 from α 1-3 galactosylation.^{213,214} One limitation is an inability to differentiate isomeric high mannose type N-glycans.²¹⁵ However, providing linkage information is an advantage of UPLC analysis of mass spectrometry, which would require a separation technique, such as ion mobility, in order to provide such information. UPLC is a well-established glycomics tool, and in many instances it is the appropriate experimental technique for N-glycan identification and analysis.

1.2.5.4 Mass Spectrometry-based Analysis of N-Glycans

As an experimental technique, mass spectrometry (MS) is well-established in the field of glycomics, and a variety of mass spectrometry techniques are well-known and widely used.²¹⁶ LC-MS is commonly done on collected fractions following LC separation of N-glycans, which can improve upon N-glycan identification from LC alone. Broadly, mass spectrometry is the most high throughput of glycomics methods, with the capacity to analyze complex mixtures much faster than LC alone. For N-glycan analysis, glycoproteins of interest are enriched and isolated from a sample, and then glycans can be either enzymatically released from associated glycoproteins or analyzed as glycopeptides.²¹⁷ Oftentimes, glycans are derivatized prior to MS analysis, with permethylation being the most common derivatization to serve to increase hydrophobicity and MS characteristics.²¹⁸ Glycans must be ionized for mass to charge detection, and typical ionization sources for glycomics mass spectrometry include electrospray ionization (ESI) and matrix-assisted laser-desorption ionization (MALDI). MALDI has the added benefit of retaining spatial information, which can be immensely valuable on tissue samples.²¹⁹ Additionally, MS/MS experiments can fragment N-glycans based on monosaccharide linkages, which allows for confident structural identification. However, without LC separation mass spectrometry alone is inadequate to distinguish between isomers. Mass spectrometry will be covered in much more detail in the following section, although it is important to note it as one of the premier glycomics experimental techniques.

1.3 Overview of Mass Spectrometry

1.3.1 Basic Principles of Mass Spectrometry

As a technology, mass spectrometry was developed in the early 20th century by J.J. Thomson primarily as a way to measure mass-to-charge ratios in the study of electrons.²²⁰ Since that time, the utilization and capabilities of mass spectrometry have expanded rapidly, with mass spectrometers found everywhere from the labs of physicists to the hospital clinical settings. Fundamentally, mass spectrometry offers the quantitative capability to measure the mass-to charge ratio (m/z) of an analyte and to record its' fragmentation pattern upon disassociation. In order to do this, mass spectrometers must include an ion source, a mass analyzer, and a detector, and all three of these elements have variability depending on the mass spectrometer. Ion sources are needed to create charged ions, as a mass spectrometer relies on ion generation in order to record m/z . Mass analyzers are used to separate these charged ions prior to reaching the detector, where mass spectra are generated. There are many additional elements depending on the type and function of a particular mass spectrometer, but these elements are recognized as critical to the basic functioning of any mass spectrometer. A detailed analysis of the nuanced operation and theory behind mass spectrometry is well outside the purview of this dissertation, which focuses on the application of the technology, but a brief synopsis of critical knowledge to the use of mass spectrometry within biomedical science is provided here.

For biological applications the two ion sources most commonly utilized are electrospray ionization (ESI) and matrix-assisted laser-desorption ionization (MALDI).

Both of these are soft ionization techniques, meaning that they induce minimal fragmentation of analytes. Malcom Dole introduced the concept of ESI in the 1960s, and it was first applied to large biological analytes by John Fenn in the 1980s.^{221,222} The concept of ESI is to inject an analyte solution into a charged capillary, which results in a fine mist at the end of the capillary of analyte-containing droplets.²²³ As the organic/aqueous solvent then evaporates, often assisted by heating, droplets shrink in size until gaseous analyte ions are generated, which can then pass on to the mass analyzer.²²⁴ ESI relies on converting analytes solubilized in a liquid solvent into gaseous ions, and can thereby be limited by analyte solubility. However, it is the preeminent ion source for analyzing liquid state biological analytes. MALDI relies on the co-crystallization of a matrix with the analyte of interest, which then absorbs UV energy of a laser to vaporize and ionize the analytes.²²⁵ The matrix is able to serve as a proton donor/receptor, allowing for both positive and negative ion mode. This ionization strategy primarily results in singly charged ions, which substantially reduces the complexity of m/z spectra. MALDI was first developed to analyze dried droplets of analytes, and the technology was then applied to mass spectrometry imaging based on the capability to generate a mass spectrum from a specific location. Through the use of a laser to ablate solid-phase analytes/matrix, MALDI allows for the retention of spatial information fundamental to mass spectrometry imaging. While they are fundamentally different in practice, the important similarity is that both ESI and MALDI ion sources result in ion generation and funnel ions towards the mass analyzer of the mass spectrometer.

The mass analyzer is responsible for the separation of ions based on their m/z so that they can be accurately measured at the detector. Mass analyzers are critical as they are responsible for mass accuracy, sensitivity, and overall resolution. There are multiple types of mass analyzers, and some of the most common include Quadrupole (Q), Time of Flight (TOF), Ion Trap, Fourier Transform (FT) analyzers.²²⁶ Quadrupole mass analyzers select ions of a particular m/z through the use of four parallel metal rods and the application of an electric field.²²⁷ Quadrupoles have become popular due to low cost, reliability, and an ability to pair with additional mass analyzers such as a Q-TOF. TOF mass analyzers rely on the acceleration of ions through a field-free flight tube, which will result in varying speeds and arrival at the detector based on mass, with smaller ions traveling through the flight tube faster.²²⁸ Newer TOF analyzers also utilize a reflectron to reflect the ions back towards the ion source before detection, which improves mass accuracy.²²⁹ TOF analyzers have become extremely popular with MALDI ion sources, and they are also popular for the capability to analyze wide m/z ranges. Ion trap analyzers are a modification to quadrupole analyzers in which a potential field is applied to trap ions within the quadrupole, which are then selectively ejected to the detector based on m/z .²³⁰ Ion traps have improved sensitivity over quadrupoles, but they have poor resolving power which has reduced their popularity over time. Fourier Transform analyzers, specifically FT-Ion Cyclotron Resonance (FT-ICR), determine m/z through the cyclotron frequency of ions in a fixed magnetic field.²³¹ FT-ICR analyzers have incredibly high resolving power and mass accuracy, which makes them a popular, albeit expensive, option.

Once ions have been separated by m/z , they must generate a signal to be recorded at the detector, which will result in mass spectra being generated. There are a multitude of available detectors for mass spectrometry, each with advantages and regular applications. An electron multiplier is among the most common of detectors, and it works through a series of dynodes, with ions striking the first dynode to generate an electron cascade to amplify and record the signal.²³² Electron multipliers are considered among the most sensitive of mass spectrometry detectors. In FT-ICR instruments, detection differs as it depends on the ICR frequency of the ion within the magnetic field.

1.3.2 Mass Spectrometry Imaging

Mass spectrometry imaging (MSI) is a well-developed application of mass spectrometry, and it has led to exciting new technological and biological developments. MSI allows for the retention of spatial information of a sample through the utilization of an ion source capable of generating analyte ions from a mixture of matrix and sample analytes, such as MALDI. Functionally, MSI relies on the generation of mass spectra at pre-determined pixel locations, commonly through the use of a laser, and these mass spectra can then be combined for the generation of heat map m/z images. Thereby, MSI allows for the untargeted analysis of many different classes of analytes, including peptides, lipids, metabolites, and glycans, without a need for sample homogenization. For biological samples, this means that the originating location of an analyte can be directly correlated to histopathological features through the overlay of histological

images and m/z images. The first mass spectrometry imaging of biological samples was done on proteins and peptides by the Caprioli laboratory in 1997, and the technology has rapidly improved and diversified since that time.²³³

Sample preparation is a critical component of successful MSI experiments, and there are basic tenets that apply regardless of the instrumentation being utilized. MSI relies on the fact that sample preparation does not disrupt the distribution of analytes in *in vivo* conditions. For biological samples, sample degradation must be halted post-dissection by either flash freezing or by formalin fixing, although formalin cross-linking is unsuitable for many analytes, with glycans being a notable exception.²³⁴ Samples then must be sectioned into thin layers (6-20 μm) and mounted onto the appropriate surface, normally a microscope slide. From this point sample preparation varies depending upon the sample, the analytes of interest, and the mass spectrometer. For N-glycan imaging, endoglycosidase(s) must be applied and incubated on the tissue surface to enzymatically free analytes of interest from associated glycoproteins.²³⁵ For MALDI mass spectrometry, a co-crystallizing matrix must be applied to allow for proper ionization within the mass spectrometer. Regardless of the analyte or the mass spectrometry application, it is critical that spatial distribution of analytes remains unaffected by sample preparation, as highly advanced MSI today can spatially resolve samples even approaching 1 μm resolution.^{236,237}

There are numerous technical applications of mass spectrometry imaging, although MALDI-IMS is the most widely utilized. MALDI requires a matrix for ionization, but there are many different choices of matrix depending on the sample/analyte, and

matrix choice and application often needs optimization for specific experiment parameters. Other MSI techniques exist including desorption electrospray ionization (DESI), and secondary ion mass spectrometry (SIMS). Both DESI and SIMS are interesting because they do not require a matrix, although they are more limited in the ionizable analytes.²³⁸ There have been numerous instrumental developments related to MSI in recent years, particularly laser-based advancements to improve spatial resolution. A lateral spatial resolution of 1.4 μm has been reported for MALDI-IMS, which is a significant improvement over spatial resolution from even 5-10 years ago, and the boundary of spatial resolution is being pushed forwards each year.²³⁹ In addition, there have been significant advances in developing separation techniques to pair with MSI, particularly through ion mobility.²⁴⁰ Another area of technical advancement has been in quantification, with new internal standard methodologies capable of much more robust quantification.²⁴¹ Taken together, the state of MSI is now in a place where it is the preeminent technology to analyze the spatial distribution of many different analytes simultaneously within a biological sample.

1.3.3 Mass Spectrometers in this Dissertation

The work in this dissertation utilized three separate mass spectrometers, all of which are MALDI-IMS instruments produced by Bruker Daltonics. The first of these mass spectrometers Bruker Solarix 7 Tesla FT-ICR MS equipped with a SmartBeam II ultraviolet laser operating at 2000 Hz with a laser spot size of 25 μm . This instrument possesses both an ESI and a MALDI source, and data collection for this dissertation

relied on the MALDI ion source. Following laser-based ionization, ions are guided into an ion beam by DC and RF voltages, and then the quadrupole selects for a specific m/z range to pass through to the hexapole. Ions are then guided into the ICR cell under high vacuum, where magnetic fields excite ions into cyclotron resonance that can be detected and transformed into mass spectra. In addition, this instrument possesses a collision cell for MS/MS validation of analyte identification through fragmentation. This mass spectrometer was the only FT-ICR instrument utilized, and thereby possessed the highest sensitivity and highest resolving power. However, the resolving power of this instrument was greater than what was required for analysis of N-glycan analytes in this work.

The TOF mass spectrometers utilized in this work included a Bruker MALDI rapifleX TissueTyper MS and a Bruker timsTOF fleX MS, with the bulk of the later work being done on the timsTOF flex. The rapifleX and timsTOF fleX both possess a SmartBeam 3D laser operating at 10,000 Hz. The timsTOF flex contains a collision cell for MS/MS analysis along with the capability for ion mobility separation, although this feature was not utilized in this dissertation. As a time-of-flight mass spectrometer, ions are accelerated by a static electric field into a field-free flight tube so that they are separated based on how smaller mass ions will achieve a higher velocity and thus reach the detector sooner. Due to this mode of ion separation and detection, acquisition speeds are much faster on both of these instruments than on the solariX FT-ICR, at the cost of decreased mass resolution. This increased acquisition speed was critical for this dissertation for the large sample sets of large liver tissue sections involved.

1.4 Broad Overview

There is a clear clinical need for improved detection of HCC, as current survival outcomes for late-stage disease are poor and unlikely to improve significantly. HCC is a serious clinical concern, particularly as incidence and mortality rates have risen in the last few decades as NAFLD incidence rises and hepatitis B/C continue to be serious threats. The largest survival benefit for HCC patients is found through early detection while the tumor is small enough to treat surgically, yet imaging modalities such as ultrasound are largely inefficient at early-stage detection. Serum biomarkers such as AFP, DCP, and AFP-L3 are all in various stages of development/clinical use, yet none have been demonstrated to be adequate replacements of ultrasound surveillance, particularly not as standalone biomarkers.

The majority of identified serum-based cancer biomarkers are glycoproteins, with some relying entirely on differential glycosylation for cancer detection. These include AFP, PSA, and CA 19-9, although there is a demonstrated need for improved performance from these markers in each of the cancer types they target. N-glycosylation has been thoroughly demonstrated to be aberrant in cancerous tissue/serum when compared to non-cancerous tissue/serum. This aberrant glycosylation depends on the cancer type but oftentimes is expressed as increased fucosylation, increased glycan branching, and differential sialylation. Mechanisms through which abnormal glycosylation drive cancer progression/metastasis are somewhat unclear, although there are well-known examples such as increased core

fucosylation driving growth factor signaling and increased branching promoting galectin binding at the cell surface.

Through literature review and previous work, it is clear that there is unlikely to be a single N-glycan structure discovered in serum with adequate sensitivity/specificity to replace current clinical HCC diagnostic tools. We hypothesized this to be due to the overall genetic/molecular heterogeneity of HCC, and thereby we aimed to analyze N-glycosylation of HCC through the lens of a genetics and clinical-based cancer subtyping system in an effort to reduce disease and glycan heterogeneity. The Hoshida HCC subtyping system classifies HCC tumors into three subgroups depending on a variety of genetic, molecular, clinical, and histological features. Utilizing this sort of information into glycan-based cancer biomarker development is novel and provides new and interesting perspective on glycan-based biomarkers, which is already one of the most exciting fields within cancer diagnostics research.

Leveraging mass spectrometry imaging into glycan-based biomarker studies is a relatively new concept and provides exciting insights into the association between glycosylation and cancer. Such an experimental technique relies on glycosidase enzymes to enzymatically cleave N-glycans without disturbing their spatial distribution. If this can be accomplished, mass spectrometry imaging allows for glycomics studies of pathologist-annotated regions of interest, as opposed to the need to homogenize tissues to do glycan analysis. Mass spectrometry also allows for the simultaneous analysis of hundreds of analytes within a given m/z window, which is ideal for N-glycan analysis. While there are shortcomings such as isomer differentiation, the experimental

technique allows for more focused cancer glycomics studies than have ever before been possible.

Chapter 2: Hypothesis

2.1 Rationale

Early detection strategies for cancer remain at the forefront of research efforts, as many curative treatments are only possible when the cancer is diagnosed at an early stage. This is particularly true for hepatocellular carcinoma (HCC), which remains difficult to detect at early stages and has a five-year survival rate of only 18%.⁵ In response to this need, there has been a focus on biomarker development, with the emergence of many glycoprotein targets that show promise to improve early-stage diagnostic sensitivity.^{8,62,67} The field of glycobiology has rapidly attracted attention for cancer biomarker development due to the important roles that the overall glycome plays in numerous biological pathways, which leads to dysregulated glycosylation in cancer progression.^{93,98,242,243} Glycosylation is a common post-translational modification of proteins, with chains of glycan subunits covalently attached to asparagine residues for N-linked glycosylation and serine/threonine residues for O-linked glycosylation. In cancer, aberrant N-glycosylation includes increased fucosylation, which has been observed in HCC along with upregulations of core fucosylation-related gene expression.^{137,244} The use of N-glycomics information in HCC provides an unique opportunity to enhance diagnostic specificity and add to the overall molecular understanding of HCC.^{72,245}

HCC is a molecularly heterogeneous disease with several disease etiologies, which makes both effective diagnosis and treatment of the disease difficult.^{16,246} In the past decade, there have been multiple efforts to classify HCC cases into genetically-based subtypes.^{40,42} These subtypes offer insight into the diversity of HCC cases, and they can

be used as a tool to advance development of clinically useful biomarkers and novel targeted therapies. Additionally, utilization of subtyped tumors in biomarker development allows for the development of markers more specific to subgroups of tumors, allowing for increased sensitivity.

Here, we propose to utilize mass spectrometry techniques to combine genetic and clinical information in a novel way with N-glycomic information in an effort to discover biomarkers for subtypes of HCC. **We hypothesize that N-glycan expression of HCC tissues/serum will correlate with the molecular and clinical differences between subtypes and thereby provide insight into the heterogenous N-glycosylation present in non-subtyped HCC tissue.**

2.2 Significance

Liver cancer is the second leading cause of cancer deaths globally, and it is a growing clinical problem due to its high mortality rate and growing incidence rate.^{1,3} This growing incidence rate is largely due to the increase in risk factors such as fatty liver disease and NASH, which is a rapidly growing problem that does not have a solution in the foreseeable future.²⁴⁷ HCC is the most common form of liver cancer, and it is the fastest rising source of cancer deaths in the United States.⁴ While the five-year survival overall is very low, the five-year survival is much higher among patients diagnosed at an eligible stage for surgical treatments, with transplantation leading to five-year survivals approaching 70%.^{8,46} This highlights the clinical need to diagnose tumors at the earliest possible stage. However, current surveillance guidelines are inadequate for the

identification and diagnosis of many early-stage tumors, particularly due to the inadequacy of ultrasound to diagnose small tumors. If a serum-based biomarker with adequate sensitivity to improve the rate of early-stage diagnoses could be identified, the survival benefit for HCC patients would be enormous. Based on a review of currently identified cancer biomarkers, N-glycomics-based analyses are most likely to yield novel HCC biomarkers to improve the detection of early-stage tumors.

2.3 Innovation

Even though the liver is the organ primarily responsible for the secretion of glycoproteins, and there is overwhelming evidence that N-linked glycosylation can be utilized for cancer biomarkers, there is still a need within the field of HCC biomarker development for more sensitive/specific markers for HCC. Disease heterogeneity has led to variable expression of many of the markers currently in use or in development for HCC, including AFP and fucosylated AFP.²⁴⁸ Therefore, it is imperative to find a way to account for this heterogeneity while identifying biomarkers that may have a synergistic effect for the purpose of improved diagnostic sensitivity. Subtyping HCC tumors into multiple classifications accounts for molecular/clinical heterogeneity of tumors, and it may account for N-glycan heterogeneity enough to improve sensitivity of markers for different HCC subtypes. **This work offers a novel N-glycomics approach by incorporating genetic/clinical data for each HCC sample analyzed in an effort to decrease overall heterogeneity of HCC N-glycosylation.** Each HCC tissue will be subtyped according to the Hoshida classification system, which incorporates both

genetic information and clinical information, and then comprehensive on-tissue and serum N-linked glycan data will be gathered. Integrating this sort of genetic information with N-glycan expression data yields a much more specific way to analyze N-glycosylation of both tissue and serum in cancer and has not previously been done. Attempting to match molecular subtypes to N-glycomic trends is also novel and will be used in an attempt to classify the heterogeneity previously seen in glycan expression on HCC tissues. If this work results in specific structures that correlate to specific subtypes, there would have significant potential for use as a biomarker for HCC. Through well-established MALDI-IMS capabilities for spatial analysis of N-glycosylation, N-glycan structures specific for tumorous regions of HCC liver tissue can be identified.^{219,234} **By combining spatial N-glycan analysis of HCC with the genetic/clinical information used to separate tumors into Hoshida subtypes, a more robust way of identifying N-glycan structures specific to subtypes of HCC tumors is developed with the goal of increasing diagnostic sensitivity overall.**

2.4 Specific Aim 1

Compare N-linked glycosylation patterns of Hoshida-subtyped hepatocellular carcinoma tissue using MALDI Imaging Mass Spectrometry

The goal of this aim is to link previously established molecular subtypes of HCC to tissue-based N-glycosylation, with the purpose of furthering the understanding of cancer-linked N-glycosylation. Hoshida HCC subtypes, developed by Dr. Yujin Hoshida,

classify HCC cases into three subtypes with unique clinical, molecular, and genetic features. Analysis of HCC, in both serum and directly on tissue, has shown overall N-glycan expression trends that are prevalent in cancer, specifically including increases in both highly branched and core fucosylated structures. However, there has been heterogeneity within HCC samples regarding these structural trends, which detracts from the potential biomarker value. We plan to utilize matrix assisted laser desorption/ionization imagine mass spectrometry (MALDI-IMS), with enzymatic release of N-linked glycans, on subtyped HCC samples in order to analyze whether this heterogeneity can be explained by known molecular differences in HCC tumors. Key breakthroughs in this aim would be 1) matching of specific N-glycan trends or individual structures directly to Hoshida subtypes of HCC 2) confirmation that HCC tumors exhibit varied N-glycosylation in a manner dependent on genetic/molecular features that can be categorized into subtypes and 3) validation of previously seen N-glycosylation patterns within HCC tissues when compared to cirrhotic or normal liver tissue.

2.5 Specific Aim 2

Correlate N-linked glycosylation of Hoshida-subtyped hepatocellular carcinoma tissues to the N-glycosylation of patient-matching serum for the development of serum-based biomarkers

The overall goal of this aim is to analyze the translation of on-tissue N-glycosylation features of HCC tumors to serum samples, which is a necessary step within

biomarker development in order to develop clinical utility. The novel aspect of this aim rests in the utilization of patient-matching serum/tissue samples, which allows for a better understanding of the progression from tumor to serum N-glycan expression while controlling for some of the demographic/lifestyle factors that may impact N-glycosylation as a whole. Additionally, the subtype-specific HCC correlation of individual serum glycoproteins will be analyzed through the utilization of a MALDI-IMS based antibody array, allowing for identification of differential N-glycosylation on particular glycoproteins of interest. Key findings in this aim would be 1) analysis of the correlation of tumor-specific N-glycosylation to serum N-glycosylation 2) furthering tissue-based N-glycan analysis with a larger sample set and 3) identification of glycoproteins of interest to distinguish cancer from non-cancer and cancer subtypes from each other.

Chapter 3: Tissue N-Glycosylation Correlates with Hepatocellular Carcinoma Subtypes

This chapter has been adapted from a manuscript published in *Molecular Cancer Research*, August 2021 19 (11): 1868-1877. ATD performed experiments, data analysis, and writing of the manuscript. AB, PA, RD, DL, and AM contributed intellectually to the manuscript. YH, AS, SL, MS, and MIF contributed to tissue sample acquisition for the manuscript.

N-glycosylation patterns correlate with hepatocellular carcinoma genetic subtypes

Andrew DelaCourt¹, Alyson Black¹, Peggi Angel¹, Richard Drake¹, Yujin Hoshida³, Amit Singal³, David Lewin², Bachir Taouli⁴, Sara Lewis⁴, Myron Schwarz⁵, M Isabel Fiel⁶ and Anand S. Mehta^{1,†}

3.1 Introduction

N-linked glycans (N-glycans), which are responsible for a wide range of physiological and cellular functions, have been demonstrated to be dysregulated in a variety of cancer types.^{93,98,101,116,151,249–253} Aberrant glycosylation has been demonstrated to contribute to tumor progression and metastasis, which has enabled research into the usage of glycan expression and glycoprotein expression as clinical biomarkers of cancer.²⁵⁴ Glycoproteins that serve as cancer biomarkers clinically include carbohydrate antigen 19-9 (CA 19-9) for pancreatic cancer and α -fetoprotein (AFP) for hepatocellular carcinoma (HCC).^{4,255} While AFP is the major tumor marker for HCC, its clinical impact in HCC screening programs is mitigated due to tumor heterogeneity, with many tumors not producing AFP. AFP alone is not recommended for surveillance given insufficient sensitivity, and the combination of abdominal ultrasound and AFP still misses over one-third of HCC cases at an early stage.⁶² Indeed, studies have identified substantial histological and molecular heterogeneity across and within HCC tumors as distinct subtypes/variants, which includes heterogeneity in AFP expression.^{256,257} In addition, a recent clinical trial showed that a subset of HCC tumors characterized by high AFP are more sensitive to a VEGFR2 antibody, ramucirumab, compared to the rest of the tumors, suggesting that biomarkers that specifically detect HCC subtypes can guide therapeutic decision making.²⁵⁸ Therefore, given its central importance to increasing curative treatment eligibility and reducing HCC-related mortality, there is a clear need for better or complimentary biomarkers to improve early HCC detection.²⁵⁹

Our lab's previous work has identified structural trends of N-glycans that are common to HCC tumors when compared to cirrhotic and healthy liver tissue.²⁴⁵ This work was done utilizing matrix-assisted laser desorption ionization imaging mass spectrometry (MALDI-IMS) to analyze N-glycan structures found on the cell surface directly on HCC tissues localized to tumor regions. Two major structural trends of glycans emerged that were correlated to HCC tumors: increased expression of fucosylated structures and increased expression of tri- and tetra-antennary branched structures. The fucosylation trends in particular support what has previously been observed in HCC serum towards biomarker discovery.^{147,244,260-262} Although it was clear that HCC tumors generally have aberrant N-linked glycosylation, there was significant inter-tumor heterogeneity of N-glycan expression, particularly in regards to the branched and fucosylated structures that were correlated to HCC regions.²⁴⁵ This heterogeneity is further supported by work that shows N-glycan heterogeneity between tumors with high AFP expression and low AFP expression.²⁴⁸ Therefore, we hypothesized that heterogeneity of N-glycosylation in HCC is correlated with specific histological, molecular, and clinical features of HCC tumors. HCC is well known to be a genetically and molecular heterogeneous disease, and recent work by a number of groups have classified HCC into specific subtypes to reduce this overall heterogeneity.^{40,42,263-265} These subtypes have been determined largely based on genome-based groupings, which have never before been linked with study of cancer N-glycomics.

In this study, we employed the latest spatial glycan profiling technology to analyze formalin-fixed paraffin-embedded (FFPE) HCC tissues from a cohort of HCC patients with the information of the Hoshida HCC molecular classification, which has been widely used as a reference system in various human HCC omics profiling consortiums, rodent models, and in vitro experimental systems.^{40,42,160,246,266–269} This classification is based on genomic features and associated with histological variants, biochemical features such as AFP, and clinical prognosis.²⁵⁶ Hoshida subtype 1 is characterized by fibrogenic molecular features such as TGF β pathway activation and accompanied with immune cell infiltrates (S1; stromal subtype). Subtype 2 is characterized by classical HCC-like features, including AFP positivity, hyper vascularity, and stemness-related cell surface markers (S2; stemness-angiogenic subtype). Subtype 3 is characterized by well differentiated histology, somatic DNA mutations in *CTNNB1* gene in half of the subtype, and less aggressive clinical tumor progression compared to S1 and S2 tumors (S3; differentiated subtype). Clinical detection of these subtypes will significantly improve prognostic prediction and enable tailored and rational treatment strategy, and diagnostic markers of these subtypes could be complementary to improve HCC diagnosis as a whole.

3.2 Materials and Methods

3.2.1 Patient Tissues and Microarrays

The tissue microarray (TMA) slide was purchased from US Biomax (Rockville, MD, Catalog Number: LV481) as unstained formalin fixed paraffin embedded (FFPE) tissue cores (5 μm thickness). The TMA contained 23 cases and 48 cores. This includes 12 HCC cases and 12 tumor-adjacent non-tumor hepatic tissue cases in duplicate. There was one matched pair between HCC cores and tumor-adjacent cores, the remainder of the HCC and tumor-adjacent samples were taken from unique patients.

FFPE tissue sections (5 μm thick) were made from 28 HCC tissue blocks from surgical resection of HCC in the background of liver cirrhosis. Samples were provided from the Icahn School of Medicine at Mount Sinai under IRB HS 13-00456 and HS 15-00888 to Dr. Yujin Hoshida and from the UT Southwestern Medical Center under IRB 102010-051 to Dr. Amit Singal. The glycan imaging work at MUSC was performed under IRB Pro00079936. Patient characteristics of these samples including sex, age, AFP levels, and etiology can be found in Table 1. HCC tissues were subtyped according to the Hoshida classification system as done previously.^{42,256} All tissues were H&E stained following MALDI-IMS analysis and tumor regions were annotated by a liver pathologist.

Variables	Discovery Cohort (n = 25)	Validation Cohort (n = 12)
Age	68 (59-72)	58 (49-68)
Male sex	18/7 (72%/28%)	7/5 (58%/42%)
Etiology (HCV/HBV/ALD/NASH/Others)	7/11/3/2/2 (28%/44%/12%/8%/8%)	4/6/1/0/1 (33%/50%/8%/0%/8%)
AFP, ng/mL	52 (7-830)	6 (4-155)
Child-Pugh class (A/B/C)	21/2/1 (88%/8%/4%)	11/1/0 (92%/8%/0%)
AJCC T stage (T1/T2/T3/T4)	8/12/2/1 (35%/52%/9%/4%)	4/6/2/0 (33%/50%/17%/0%)
Hoshida HCC subtype (S1/S2/S3.1/S3.2)	6/4/4/11 (24%/16%/16%/44%)	5/2/1/4 (42%/17%/8%/33%)
Categorical variables are shown as n (%). Continuous variables are shown as median (IQR).		

Table 1. Patient Demographics. All tissues between the discovery and validation cohorts are included. Abbreviations: HCV, hepatitis C virus; HBV, hepatitis B virus; ALD, alcohol-related liver disease; NASH, non-alcoholic steatohepatitis; AFP, alpha-fetoprotein; AJCC, American Joint Committee of Cancer.

3.2.2 FFPE Tissue Preparation for MALDI-IMS

HPLC grade methanol (Catalog No. A452SK-4), ethanol (Catalog No. 22-032-601), acetonitrile (Catalog No. A998-1), xylene (Catalog No. X3S-4), and water (Catalog No. W5-1) were obtained from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid (Catalog No. W302031) and α -cyano-4-hydroxycinnamic acid (Catalog No. C89892) were obtained from Sigma-Aldrich (St. Louis, MO). Peptide-N-glycosidase F (PNGase F) Prime was cloned, expressed, and purified in-house as previously described.²¹⁹

FFPE tissues and the TMA were prepared according to a previously published protocol.^{219,234,270} Tissue Tack microscope slides were purchased from Polysciences Inc (Warrington, PA, Catalog No. 24216), and indium tin oxide glass slides were purchased from Delta Technologies (Loveland, CO, Catalog No. CB-40IN-S111). Briefly, slides were

heated to 60°C for 1 hour, washed with xylenes, and rehydrated with a series of ethanol and water washes. Slides were then processed by antigen retrieval, heating to 95° C for 30 minutes in a Decloaking Chamber in a 10 mM citraconic anhydride buffer, pH 3. Slides were cooled to room temperature and buffer exchange was done to replace buffer with 100% water. PNGase F Prime™ was applied using a M5 TM-Sprayer Tissue MALDI Sample Preparation System (HTX Technologies, LLC). After a two-hour incubation at 37°C, MALDI matrix α -cyano-4-hydroxycinnamic acid (0.042 g CHCA in 6 mL 50% acetonitrile/49.9% water/0.1% TFA) was sprayed by the M5 TM-Sprayer. The full workflow to prepare FFPE tissues for MALDI N-glycan imaging is shown in Figure 5.

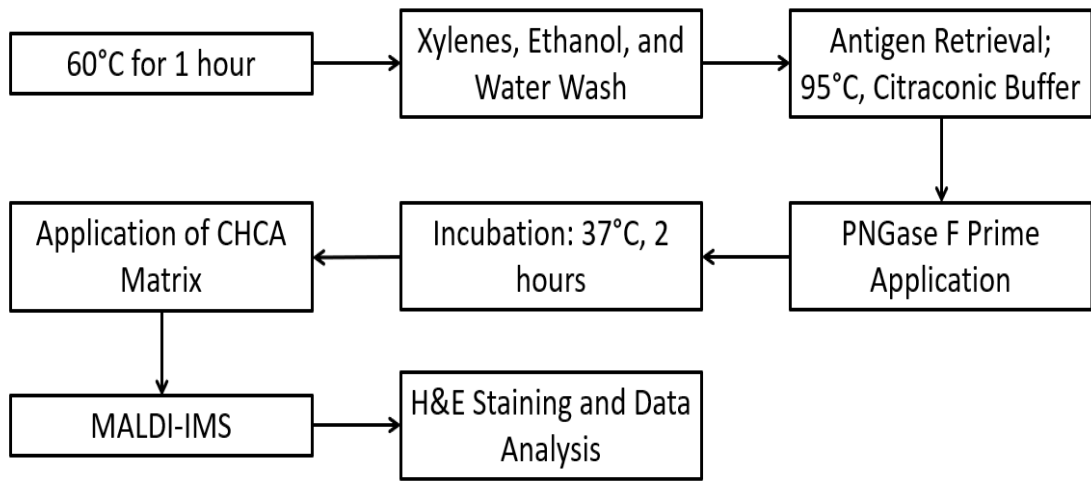


Figure 5. MALDI-IMS Tissue Preparation Workflow. A flowchart depicting the workflow to prepare FFPE tissues from dewaxing through MALDI-IMS and data analysis.

3.2.3 MALDI-IMS of N-Glycans

Slides were imaged on a MALDI FT-ICR (Solarix Legacy 7T, Bruker Daltonics) mass spectrometer in positive ion, broadband mode (m/z 500-5000) and a MALDI TOF (RapifleX TissueTyper, Bruker Daltonics) in positive ion mode (m/z 600-3500). Images were collected on the FT-ICR at a 150 μM raster with 200 laser shots per pixel. Images were collected on the MALDI-TOF at a 50 μM raster with 200 laser shots per pixel. Data was visualized and analyzed using FlexImaging 5.0 and SCiLS Lab 2019c (Bruker). Peaks were assigned to N-glycan structures based upon mass using a database of N-glycan structures built with consideration to biosynthetic pathways.²⁷¹ Putative structures are shown based on previous databases through use of databases built with use of GlycoWorkBench (RRID:SCR_000782).^{127,219,234} Glycan structures, m/z values, and mass error can be found Table 2 for FT-ICR data and Table 5 for MALDI-TOF data.

3.2.4 Statistical Analysis

After MALDI-IMS analysis, statistical analyses were done in order to evaluate differences between HCC tissue and adjacent non-HCC tissue in both the TMA and whole tissue analyses. For the TMA, area under the peak (AUP) was determined using SCiLS for each m/z value in each tissue core, and the average AUP was determined for each m/z of both HCC cores and non-HCC cores. Unpaired student's *t* tests were used to determine glycan structures that were significantly increased in either HCC or non-HCC cores, with a cutoff of $p < 0.05$. For the whole tissues, HCC and adjacent liver tissue regions were annotated by a pathologist. For the cohort analyzed via FT-ICR, AUP was

determined for each m/z value for both the tumor region and adjacent non-tumor (primarily cirrhotic) region. An AUP fold change of >3 signified a tumor-associated increase to glycan expression. Significant differences between groups were determined using a Wilcoxon Rank-Sum Test, with a cutoff of $p < 0.05$. For the cohort analyzed via MALDI TOF, SCiLS spatial co-localization feature was utilized, which identifies m/z peaks with resultant images that demonstrate a spatial correlation in signal intensity to the tumor region, with a threshold of a >0.4 Pearson coefficient to determine positive correlation. A top-hat baseline correction through SCiLS was utilized for the MALDI-TOF mass spectrometry imaging runs.

3.3 Results

3.3.1 Inter-Tumor Heterogeneity of N-Glycan Expression in HCC Tissues

Previous work has suggested that there are distinguishable differences in N-glycan expression in both HCC serum and HCC tissues when compared to both cirrhotic and healthy liver tissue.^{89,245,272} To validate these results, we analyzed the N-glycome of an independent HCC tissue microarray (TMA) through MALDI-IMS FT-ICR (Figure 6). The TMA was H&E stained to compare to MALDI images, which is shown in Figure 6A. Tissue preparation for N-glycan analysis through MALDI-IMS involves the usage of the enzyme PNGase F Prime™, which cleaves N-glycans from arginine residues of glycoproteins while retaining spatial localization of the glycan. In this analysis, 88 unique glycan structures were observed, with the area under the curve (AUP) of each m/z peak analyzed for each core (Table 2). Peaks were selected based on the known theoretical m/z values from a database of N-glycan structures. Twenty-three of these structures were significantly increased in the HCC cores, and seven of the structures were significantly increased in the tumor-adjacent non-tumor cores (Tables 3 and 4). Of the 23 structures that were significantly increased in HCC samples, 19 were fucosylated and 16 structures were complex glycans that were either bisecting, tri-antennary, and tetra-antennary structures (Figure 6B and 6C). Additionally, none of the seven glycans that are increased in tumor-adjacent cores were fucosylated and instead are primarily high mannose or non-fucosylated biantennary structures. A representative image of one of these glycans increased in non-tumor cores is shown in Figure 6D. Fucosylated and/or branched glycans made up the majority of glycans that were increased in HCC cores, but

there were a few structures that did not fit this pattern. Some of the glycans increased in HCC cores were relatively simple structures that were non-fucosylated and/or simple structures with only two antennae, an example of which is seen in Figure 6E.

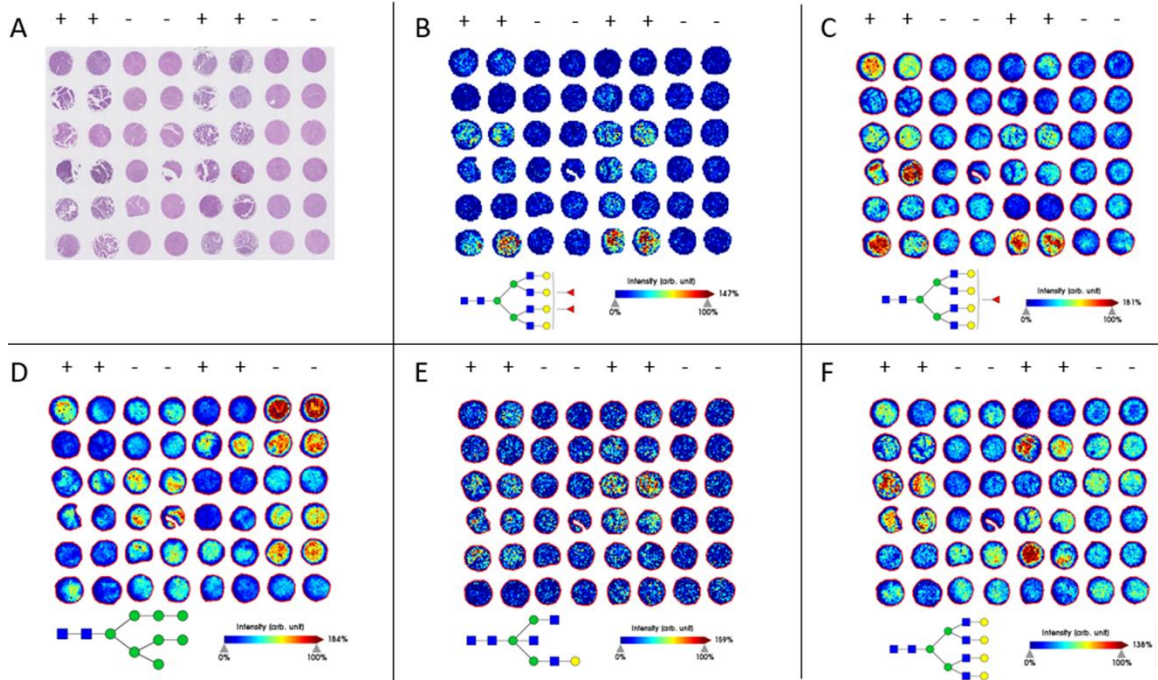


Figure 6. TMA Analysis of the N-glycosylation of HCC and Normal Liver. (A): H&E staining of the TMA. The organization of the TMA includes 2 cores each from 12 unique HCC tumors, and 2 cores each from 12 unique normal liver tissue adjacent to HCC cases. The columns indicated with a “+” include horizontally paired HCC cores, and the columns indicated with a “-” include horizontally paired normal adjacent cores. (B-F): MALDI-IMS images of representative glycan structures. (B-C): Structures that are significantly enriched in HCC cores. $m/z= 2685.9551$ (B); $m/z= 2539.9045$ (C); (D): A structure that is more abundant in normal adjacent cores than HCC cores. $m/z= 1743.5816$ (E): A structure that is significantly enriched in HCC tumors but is observed at relatively low levels. $m/z= 1704.6180$ (F): A structure that is not significantly enriched in HCC cores as a whole, but clearly over-abundant in select tumors. $m/z= 2393.8436$

Observed <i>m/z</i>	Theoretical <i>m/z</i>	Error in PPM	Glycan Structure
771.2694	771.2822	16.6346	Hex2HexNAc2
933.317	933.3170	0.0000	Hex3HexNAc2
1079.3753	1079.3749	0.3706	Hex3dHex1HexNAc2
1095.3690	1095.3698	0.7303	Hex4HexNAc2
1136.3952	1136.3964	1.0560	Hex3HexNAc3
1257.4225	1257.4226	0.0795	Hex5HexNAc2
1282.4522	1282.4543	1.6375	Hex3dHex1HexNAc3
1298.4487	1298.4492	0.3851	Hex4HexNAc3
1339.4736	1339.4758	1.6424	Hex3HexNAc4
1419.4753	1419.4754	0.0704	Hex6HexNAc2
1444.5071	1444.5071	0.0000	Hex4dHex1HexNAc3
1460.5029	1460.502	0.6162	Hex5HexNAc3
1485.5326	1485.5337	0.7405	Hex3dHex1HexNAc4
1501.5309	1501.5286	1.5318	Hex4HexNAc4
1542.5564	1542.5551	0.8428	Hex3HexNAc5
1581.5279	1581.5282	7.4882	Hex7HexNAc2
1606.5714	1606.5599	0.1897	Hex5dHex1HexNAc3
1611.6212	1611.5266	7.1582	Hex4HexNAc3NeuAc1+ 2Na
1622.5582	1622.5548	4.4055	Hex6HexNAc3
1647.5873	1647.5865	2.0955	Hex4dHex1HexNAc4
1663.5835	1663.5814	0.4856	Hex5HexNAc4
1688.6124	1688.613	0.3553	Hex3dHex1HexNAc5
1704.6180	1704.6079	5.9251	Hex4HexNAc5
1743.5816	1743.581	0.3441	Hex8HexNAc2
1809.6404	1809.6393	0.6079	Hex5dHex1HexNAc4
1825.6276	1825.6342	3.6152	Hex6HexNAc4
1850.6643	1850.6659	0.8646	Hex4dHex1HexNAc5
1866.6591	1866.6608	0.9107	Hex5HexNAc5
1905.6347	1905.6338	0.4723	Hex9HexNAc2
1911.5935	1911.5859	3.9758	Hex5dHex1HexNAc4 + 2Na + SO4
1954.6746	1954.6768	1.1255	Hex5HexNAc4NeuAc1
1955.6802	1955.6972	8.6926	Hex5dHex2HexNAc4
1976.6593	1976.6666	3.6931	Hex5HexNAc4NeuAc1 + 2Na
1992.6459	1992.6537	3.9144	Hex5HexNAc4NeuGc1 + 2Na
2010.6535	2010.6224	15.4678	Hex7dHex1HexNAc3 + SO4
2012.7208	2012.7187	1.0434	Hex5dHex1HexNAc5
2028.7122	2028.7136	0.6901	Hex6HexNAc5
2057.6575	2057.6439	6.6095	Hex5dHex2HexNAc4 + 2Na + SO4
2067.6925	2067.6837	4.2560	Hex10HexNAc2
2069.7194	2069.7402	10.0491	Hex5HexNAc6
2100.7356	2100.7347	0.4284	Hex5dHex1HexNAc4NeuAc1
2101.7371	2101.7551	8.5643	Hex5dHex3HexNAc4
2122.7149	2116.7296	4.5225	Hex5dHex1HexNAc4NeuAc1 + 2Na
2157.7729	2157.7562	7.7400	Hex5HexNAc5NeuAc1
2158.7790	2158.7766	1.1117	Hex5dHex2HexNAc5

2163.7230	2163.7432	3.6511	Hex4dHex1HexNAc5NeuAc1 + 2Na
2174.7784	2174.7715	3.1727	Hex6dHex1HexNAc5
2179.7457	2179.7460	0.1230	Hex5HexNAc5NeuAc1 + 2Na
2215.7670	2215.798	16.8793	Hex5dHex1HexNAc6
2245.7753	2245.7722	1.3804	Hex5HexNAc4NeuAc2
2246.7814	2246.7926	4.9844	Hex5dHex2HexNAc4NeuAc1
2267.7479	2267.7620	6.2176	Hex5HexNAc4NeuAc2 + 2Na
2268.7737	2268.7824	3.8206	Hex5dHex2HexNAc4NeuAc1 + 2Na
2276.7054	2276.7180	5.5343	Hex6dHex1HexNAc5 + 1SO4 + 2Na
2289.7335	2289.7326	0.3931	Hex5HexNAc4NeuAc2 + 3Na
2303.8154	2303.8141	0.5643	Hex5dHex1HexNAc5NeuAc1
2304.8300	2304.8345	1.9524	Hex5dHex3HexNAc5
2319.8138	2319.8090	2.0691	Hex6HexNAc5NeuAc1
2320.8276	2320.8294	0.7756	Hex6dHex2HexNAc5
2325.7902	2333.6647	5.8904	Hex5dHex1HexNAc5NeuAc1 + 2Na
2333.6991	2333.6647	14.7408	Hex5HexNAc5NeuGc1 + 2SO4
2341.7926	2341.7988	2.6475	Hex6HexNAc5NeuAc1 + 2Na
2377.8445	2377.8509	2.6915	Hex6dHex1HexNAc6
2391.8300	2391.8301	0.0418	Hex5dHex1HexNAc4NeuAc2
2393.8436	2393.8458	0.9190	Hex7HexNAc6
2413.8292	2413.8199	3.8528	Hex5dHex1HexNAc4NeuAc2 + 2Na
2421.8446	2421.7984	19.0767	Hex9HexNAc3NeuAc1 + 2Na
2435.8095	2465.8669	0.0821	Hex5dHex1HexNAc4NeuAc2 + 3Na
2465.8649	2465.8669	0.8111	Hex6dHex1HexNAc5NeuAc1
2487.8475	2487.8567	3.6980	Hex6dHex1HexNAc5NeuAc1 + 2Na
2539.9045	2539.9037	0.3150	Hex7dHex1HexNAc6
2632.9040	2632.8942	3.7343	Hex6HexNAc5NeuAc2 + 2Na
2633.9218	2633.9146	2.7336	Hex6dHex2HexNAc5NeuAc1 + 2Na
2653.9568	2653.9718	5.6519	Hex5dHex4HexNAc6
2684.9366	2684.9412	1.7133	Hex7HexNAc6NeuAc1
2685.9551	2685.9616	2.4200	Hex7dHex2HexNAc6
2706.9144	2706.9310	6.1324	Hex7HexNAc6NeuAc1 + 2Na
2742.9912	2742.9831	2.9530	Hex7dHex1HexNAc7
2758.9686	2758.978	3.4071	Hex8HexNAc7
2779.9528	2779.9725	7.0864	Hex6dHex3HexNAc5NeuAc1 + 2Na
2830.9227	2830.9991	26.9869	Hex7dHex1HexNAc6NeuAc1
2832.0145	2832.0195	1.7655	Hex7dHex3HexNAc6
2853.0016	2852.9889	4.4515	Hex7dHex1HexNAc6NeuAc1 + 2Na
2905.0552	2905.0359	6.6436	Hex8dHex1HexNAc7
3124.1081	3124.1102	0.6722	Hex9HexNAc8
3218.1263	3218.1211	1.6258	Hex8dHex1HexNAc7NeuAc1 + 2Na
3270.1746	3270.1681	1.9877	Hex9dHex1HexNAc8
3635.3285	3635.3003	7.7573	Hex10dHex1HexNAc9

Table 2: N-Glycans Analyzed on the MALDI FT-ICR. Included is the full *m/z* peak list of identified N-glycans on the MALDI FT-ICR mass spectrometer.

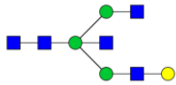

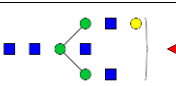

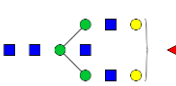
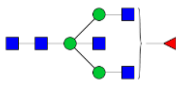

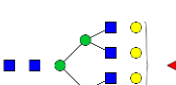



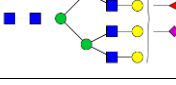
Observed <i>m/z</i>	Proposed Glycan Structure	Composition	P value
1704.5736		Hex4HexNAc5	1.0874E-06
2377.8183		Hex6dHex1HexNAc6	4.3991E-06
1850.6241		Hex4dHex1HexNAc5	6.7922E-06
2685.9074		Hex7dHex2HexNAc6	8.4026E-05
2012.6801		Hex5dHex1HexNAc5	0.0001119
1688.5748		Hex3dHex1HexNAc5	0.0002022
2831.9725		Hex7dHex3HexNAc6	0.0002513
2539.8637		Hex7dHex1HexNAc6	0.0005301
1647.5597		Hex4dHex1HexNAc4	0.0005561
2174.7385		Hex6dHex1HexNAc5	0.001602
2320.7831		Hex6dHex2HexNAc5	0.002175
2830.9227		Hex7dHex1HexNAc6NeuAc1	0.002914

Table 3. Significantly HCC-associated N-glycans within TMA. Included are the 12 N-glycan structures that are most significantly associated with HCC cores of the TMA, out of 23 total that were significantly HCC-associated.


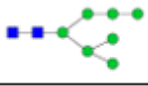
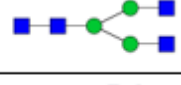




Observed m/z	Proposed Glycan Structure	Composition	P value
1743.5816		Hex8HexNAc2	1.3848E-05
1581.5279		Hex7HexNAc2	2.6587E-05
1339.4736		Hex3HexNAc4	0.0008810
2758.9686		Hex8HexNAc7	0.004078
1136.3952		Hex3HexNAc3	0.004984
2267.7479		Hex5HexNAc4NeuAc2 + 2Na	0.03492
1419.4753		Hex6HexNAc2	0.03505

Table 4. Significantly Normal-Associated N-glycans within TMA. Included are the 7 N-glycan structures that are most significantly associated with normal cores of the TMA.

While many of the N-glycan expression trends that have been previously identified are validated here, inter-tumor and intra-tumor heterogeneity remains. This sort of heterogeneity is clear throughout all of the *m/z* images shown in Figure 6, where there are no glycans that are consistent in expression throughout all of the HCC cores or throughout all of the non-tumor tumor-adjacent cores. This sort of heterogeneity is highlighted in Figure 6F, which is a tetra-antennary glycan structure that has previously been shown to be often overexpressed in HCC tissues. This glycan was not significantly overexpressed in this TMA, with a p-value of 0.125, yet several of the HCC cases within

the TMA express this glycan at very high levels compared to the non-HCC cores. Even the glycans that are significantly increased in HCC cores of this TMA vary substantially from tumor to tumor in expression (Figure 6 B-C). N-glycosylation of HCC tissue is distinct from non-HCC tissue, but there are clearly additional factors beyond tumor presence that account for which structures are increased in expression in a tumor and to what degree they are increased in expression.

3.3.2 Tumor-Specific N-Glycan Expression Associated with HCC Molecular Subtypes

In order to analyze N-glycosylation of HCC samples while accounting for genetic, molecular, and clinical variation, we subsequently analyzed HCC tumors that were classified using the Hoshida classification system.^{42,263} A direct comparison was made between each HCC tumor itself and its surrounding adjacent tissue for each sample, which was evaluated as the fold change for each glycan between regions. The adjacent tissue that was utilized in the analysis consisted primarily of cirrhotic/non-tumor liver tissue, with necrotic and fibrotic regions excluded. In total, thirty-seven subtyped tissues were examined in two defined sets. The first set consisted of twenty-five samples (six S1, four S2, fifteen S3 tumors) and was examined on the FT-ICR mass spectrometer as a part of the “discovery cohort”. Subsequently, the remaining twelve samples were examined on the MALDI-TOF instrument as a “validating cohort”. The discovery cohort was analyzed on a high mass accuracy and high sensitivity mass spectrometer in order to find specific N-glycosylation structural features associated with molecular subtypes. The validating cohort was analyzed on a more clinically accessible MALDI-TOF

instrument at a higher spatial resolution to both evaluate the clinical applicability of the subtype-linked N-glycosylation and to further observe potential intra-tumor heterogeneity. These samples were utilized to represent the diversity of genetic characteristics of HCC tumors and not necessarily to mirror demographic characteristics found in the clinic such as etiology or gender, which have not been previously shown to be correlated to N-glycan expression.

As seen in Figure 7, the discovery cohort showed that all three subtypes demonstrate altered N-glycosylation of HCC tumors when compared to surrounding cirrhotic/non-tumor tissue, which further validates what previous work and the TMA analysis demonstrated broadly. For this work, an area under the peak fold change of three or higher from adjacent tissue to the tumor region was considered to be a tumor-associated increased expression of a glycan structure. Glycan structures varied in the magnitude of the fold change/tumor-associated increase, but of particular interest is the number of unique glycan structures that are increased within tumors, which implies broad dysregulation over the expression of N-glycan structures.

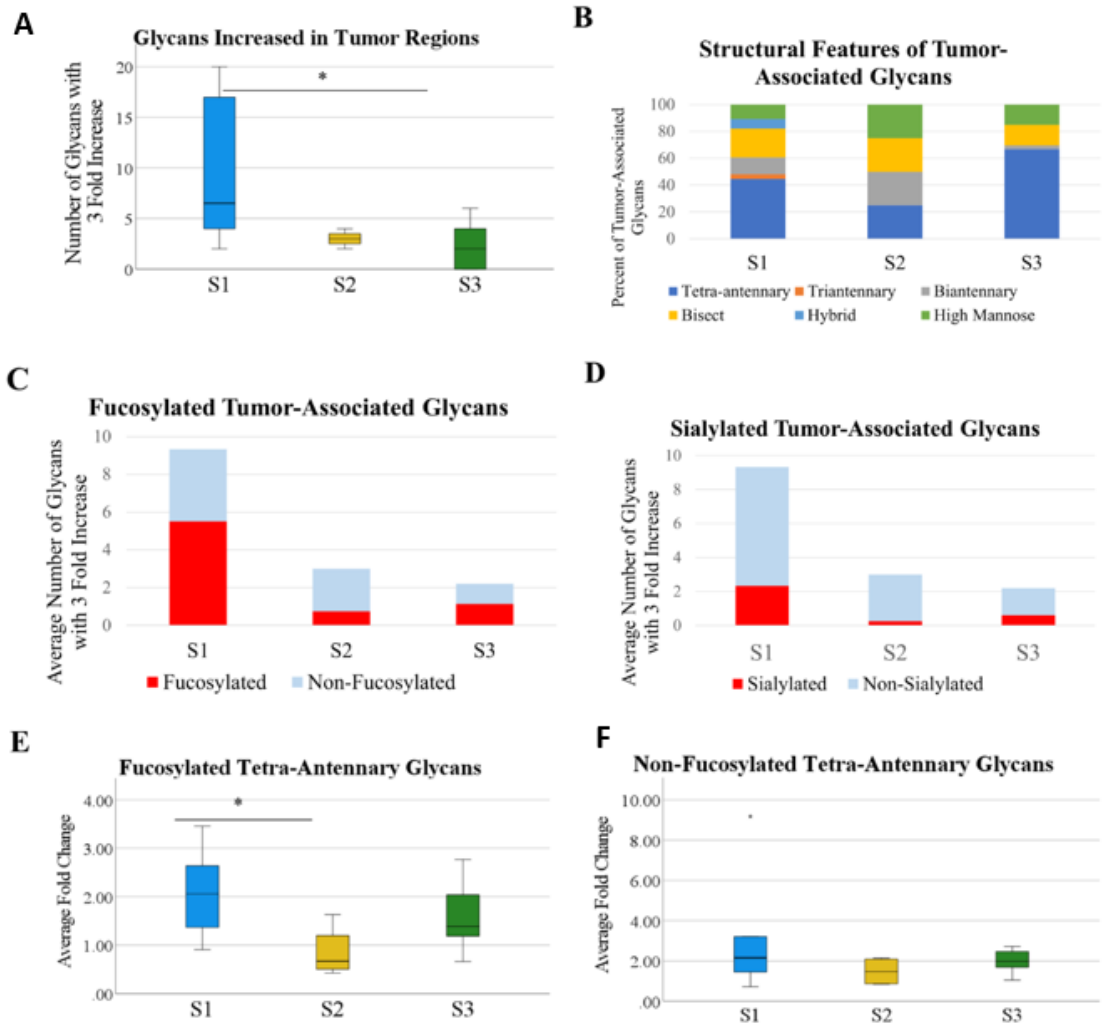


Figure 7. N-Glycosylation of HCC Subtypes. (A): Analysis of the number of unique glycan structures with a 3+ fold area under the peak (AUP) increase in the tumor when compared to adjacent normal/cirrhotic tissue. n=5 (S1), n=4 (S2), n=15 (S3). Statistical test includes Wilcoxon Rank-Sum test (* = $p < 0.05$) (B): The structural features of glycans that are increased in the tumor for each subtype. Structural features are determined based upon proposed structures for m/z values. (C): Analysis into the increase of fucosylated structures in tumors of each subtype. (D): Analysis into the increase of sialylated structures for tumors of each subtype. (E): The average fold change of all fucosylated tetra-antennary glycans by subtype. Average fold change was calculated through the sum AUP of these glycans in each tumor and non-tumor region. Statistical test includes Wilcoxon Rank-Sum test (* = $p < 0.05$). (F): The average fold change of all non-fucosylated tetra-antennary glycans by subtype. Statistical test includes Wilcoxon Rank-Sum test (* = $p < 0.05$).

All three subtypes demonstrate differentiating glycans from the non-tumor tissue, which vary based on the tumor, but S1 tumors in this cohort had significantly more unique glycan structures that were increased in the tumor region than S2/S3 per tumor (Figure 7A). When accounting for structural features of these glycans, the glycans that are increased in the tumor are largely made up of bisecting and tetra-antennary glycans in all three subtypes, which aligns with previously held ideas that branching is generally increased in HCC tumors (Figure 7B). When comparing fucosylation patterns between the subtypes, S1 tumors overexpress fucosylated structures with higher frequency than S2 or S3 tumors (Figure 7C). Increased expression of fucosylated structures appears to be a feature in S3 tumors similarly to S1, but there are significantly fewer overexpressed structures overall in S3 tumors. However, overexpression of fucosylated structures does not appear to occur at the same level in S2 tumors compared to S1 tumors, which offers a potential distinguishing feature of the two clinically aggressive subtypes. It does not appear as though a major share of tumor-associated glycans are sialylated in any of the three subtypes, which aligns with previous work that fucosylation is more relevant to HCC than sialylation (Figure 7D).

Fucosylated tetra-antennary structures have been previously shown to be correlated with HCC in general, although lacking the necessary sensitivity to serve as a standalone biomarker. However, fucosylated tetra-antennary glycans are increased at significantly higher levels in S1 tumors than S2 tumors, while there is no difference between the three subtypes in the increased expression of non-fucosylated tetra-antennary structures (Figure 7E/F). This suggests that tetra-antennary structures are

increased in expression in all subtypes of HCC, but fucosylated tetra-antennary structures vary in expression between subtypes of HCC, with higher expression in the clinically aggressive S1 tumors than the clinically aggressive S2 tumors.

The primary observed feature of the N-glycosylation in S1 tumors is the increased expression of fucosylated and/or branched glycan structures. In Figure 8, representative images from three of the S1 tumors are shown, with the tumors outlined in blue on the H&E stains (Figure 8A-C). In each of these three tumors, there are many tumor-increased glycan structures which fit the themes of increased branching and increased fucosylation. The expression of both a tetra-antennary structure with a single fucose (Figure 8D-F) and a tri-antennary structure with a single fucose (Figure 8G-I) are increased in all three of these tumors. These individual glycan structures are not increased in all six of the S1 tumors in this cohort, but these branched and fucosylated glycans represent the type of structures that are often increased in S1 tumors. These *m/z* images also demonstrate that HCC tumors contain intra-tumor heterogeneity, which was expected but requires further analysis in order to evaluate.

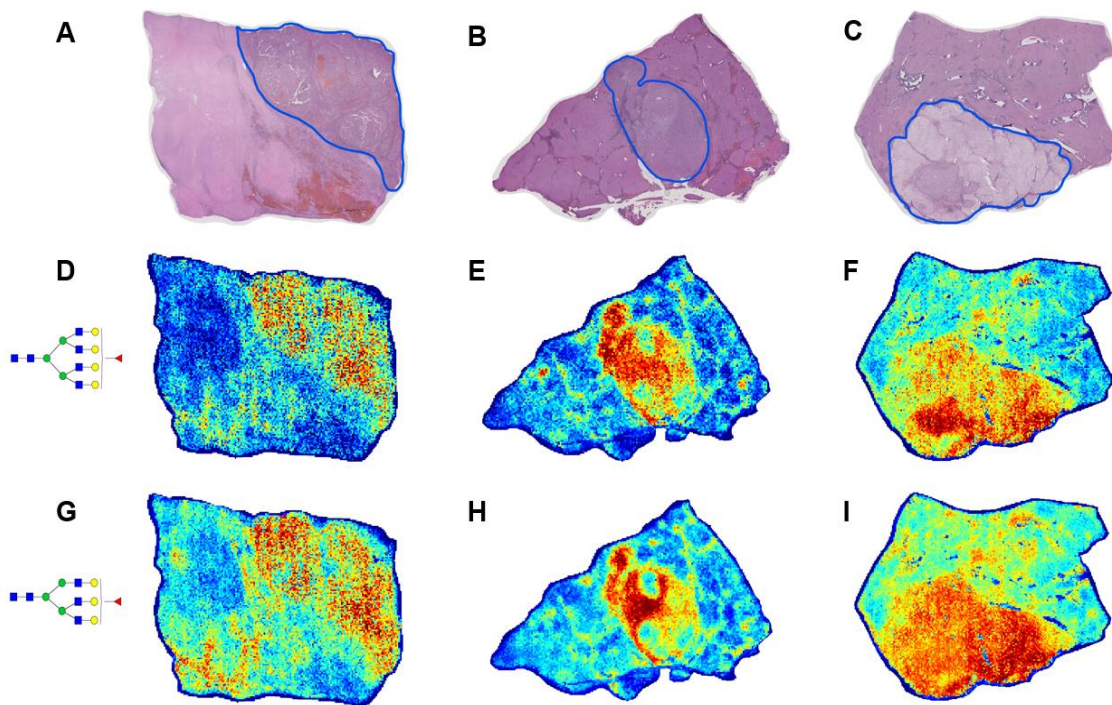


Figure 8. Representative S1 Tumors. (A-C): H&E stains of three S1 tissues. Tumors are outlined in blue. Regions of the adjacent tissue with significant necrosis or fibrosis were not included in glycosylation analyses. (D-F): Expression of a tetra-antennary structure with a single fucose, $m/z= 2539.9045$ (G-I): Expression of a tri-antennary structure with a single fucose, $m/z= 2174.7784$

Fucosylation in S2 tumors is of particular interest, as there is a clear difference in fucosylation patterns between S1 and S2 tumors as demonstrated in Figure 9. As seen previously, fucosylated structures being increased within the tumor is not a feature in S2 tumors. Additionally, a large share of the structures with a three-fold increase in the adjacent tissue compared to the tumor are fucosylated (Figure 9A). This would suggest that there is actually decreased expression of fucosylated structures in S2 tumors compared to adjacent non-HCC liver tissue, which directly opposes what is seen in S1 tumors. Representative images demonstrating this fucosylation pattern are shown in Figure 9B, where a simple biantennary structure is tumor-correlated, while the fucosylated structures shown to be tumor-correlated in S1 tumors are expressed more in the adjacent non-tumor liver tissue than in this S2 tumor. An analysis of individual glycan structures reveals that fucosylated structures that are frequently tumor-correlated in S1 tumors are expressed at significantly higher levels in S1 tumors than in S2 tumors regardless of the adjacent tissue. (Figure 9C). These four fucosylated structures are some of the most abundant tri-antennary, bisecting, and tetra-antennary fucosylated structures observed in these tissues, which suggests that fucosylation patterns are distinguishably different between these subtypes. Overall, fucosylation appears to be directly involved in tumor progression of many S1 tumors, while the expression of fucosylated structures is decreased in S2 tumors.

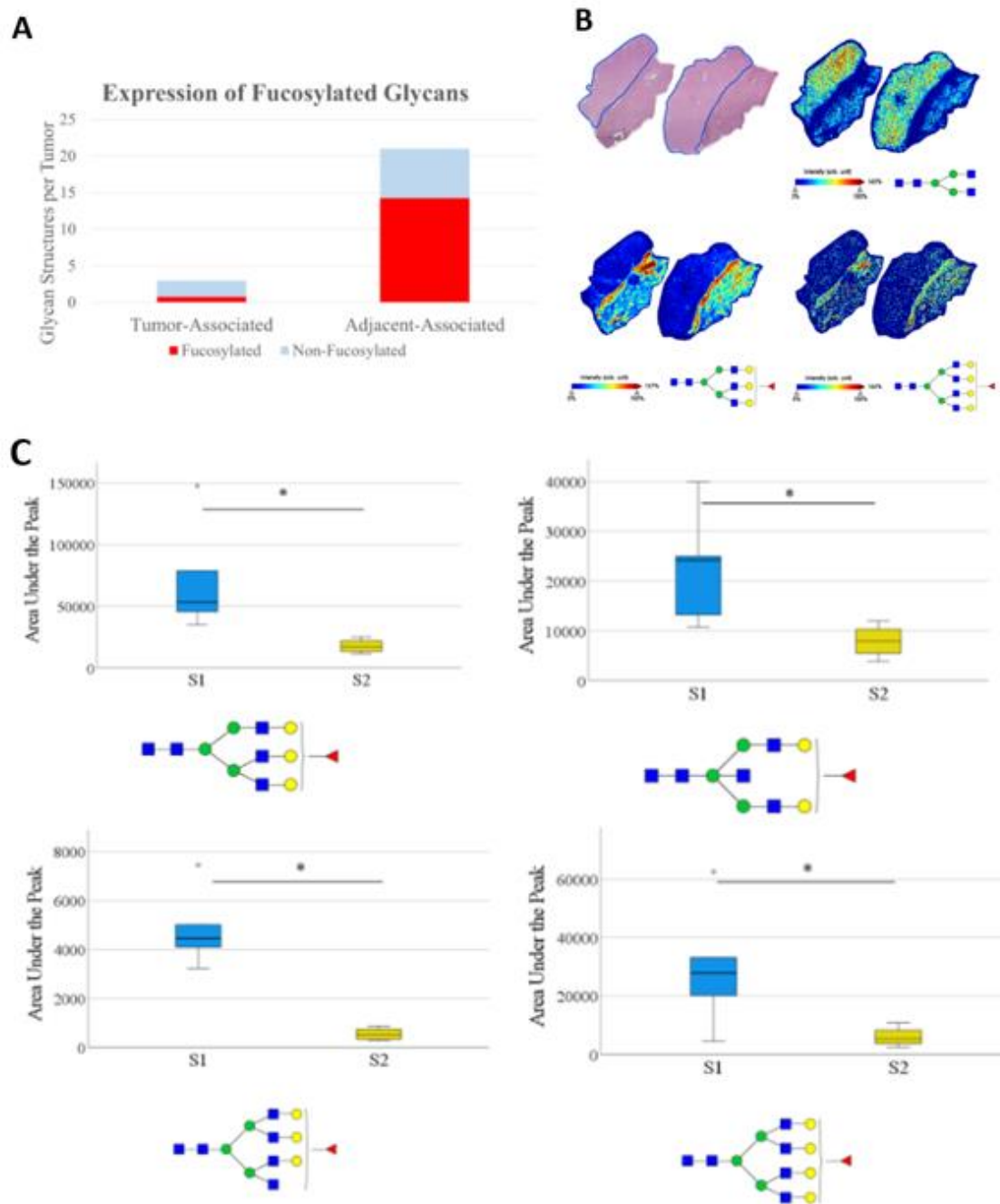


Figure 9. Fucosylation of Subtype 2 Tumors. (A): Number of fucosylated and non-fucosylated glycans in HCC and non-HCC adjacent tissue with a 3+ fold change. (B): Representative images of an S2 tissue. Tumor outlined in blue, with normal adjacent liver tissue. $m/z= 1339.4736$ (top right), $m/z= 2174.7784$ (bottom left), $m/z= 2539.9045$ (bottom right). (C): AUP of four glycans for all S1 tumors and S2 tumors, without inclusion of adjacent tissue. Four fucosylated glycans are significantly increased in S1 tumors, $m/z= 2174.7784$, 2012.7208 , 2377.8445 , 2539.9045 (top left, top right, bottom left, bottom right). Statistical test includes Wilcoxon Rank-Sum test (* = $p < 0.05$).

3.3.3 Validation of Differentiation of Subtypes by N-Glycosylation through MALDI-TOF

To confirm subtype-based glycosylation trends observed in the initial cohort of HCC tissues, an additional cohort of 12 subtyped HCC tumors was analyzed using a MALDI-TOF instrument, with the list of detected glycans listed in Table 5. Specifically, the aim of this cohort is to validate the distinction of S1 tumors from S2 and S3 tumors based upon an increased expression of fucosylated N-glycans. TOF-based mass spectrometry involves currently more clinically applicable instrumentation, albeit with a decreased sensitivity, which is another reason to validate findings on a more accessible instrument. Data analysis was done by determining m/z peaks determined to be spatially correlated to the tumor regions by Pearson coefficient, which differed from analysis of the discovery cohort due to variations in signal intensity from run to run on this instrument.

Observed <i>m/z</i>	Theoretical <i>m/z</i>	Error in PPM	Glycan Structure
933.318	933.317	1.0714	Hex3HexNAc2
1079.371	1079.375	3.7058	Hex3dHex1HexNAc2
1095.370	1095.370	0.0000	Hex4HexNAc2
1136.395	1136.396	0.8800	Hex3HexNAc3
1257.423	1257.423	0.0000	Hex5HexNAc2
1282.446	1282.454	6.2380	Hex3dHex1HexNAc3
1298.446	1298.449	2.3104	Hex4HexNAc3
1339.473	1339.476	2.2397	Hex3HexNAc4
1419.475	1419.475	0.0000	Hex6HexNAc2
1444.495	1444.507	8.3073	Hex4dHex1HexNAc3
1460.513	1460.502	7.5317	Hex5HexNAc3
1485.531	1485.534	2.0195	Hex3dHex1HexNAc4
1501.528	1501.529	0.6660	Hex4HexNAc4
1542.547	1542.555	5.1862	Hex3HexNAc5
1581.526	1581.528	1.2646	Hex7HexNAc2
1606.585	1606.560	15.5612	Hex5dHex1HexNAc3
1622.543	1622.555	7.3957	Hex6HexNAc3
1647.596	1647.587	5.4625	Hex4dHex1HexNAc4
1663.582	1663.581	0.6011	Hex5HexNAc4
1688.603	1688.613	5.9220	Hex3dHex1HexNAc5
1704.597	1704.608	6.4531	Hex4HexNAc5
1743.582	1743.581	0.5735	Hex8HexNAc2
1792.627	1792.624	1.6735	Hex4HexNAc4NeuAc1
1809.641	1809.639	1.1052	Hex5dHex1HexNAc4
1825.626	1825.634	4.3820	Hex6HexNAc4
1850.660	1850.666	3.2421	Hex4dHex1HexNAc5
1866.669	1866.661	4.2857	Hex5HexNAc5
1905.636	1905.634	1.0495	Hex9HexNAc2
1911.567	1911.586	9.9394	Hex5dHex1HexNAc4 + 2Na + SO4
1955.699	1955.697	1.0227	Hex5dHex2HexNAc4
1976.671	1976.667	2.0236	Hex5HexNAc4NeuAc1 + 2Na
2010.670	2010.622	23.8732	Hex7dHex1HexNAc3 + SO4
2012.719	2012.719	0.0000	Hex5dHex1HexNAc5
2028.712	2028.714	0.9858	Hex6HexNAc5
2057.702	2057.644	28.1876	Hex5dHex2HexNAc4 + 2Na + SO4
2067.680	2067.684	1.9345	Hex10HexNAc2
2122.738	2122.725	6.1242	Hex5dHex1HexNAc4NeuAc1 + 2Na
2174.767	2174.772	2.2991	Hex6dHex1HexNAc5
2215.794	2215.730	28.8844	Hex5dHex1HexNAc6
2320.835	2320.829	2.5853	Hex6dHex2HexNAc5
2341.809	2341.799	4.2702	Hex6HexNAc5NeuAc1 + 2Na
2377.842	2377.851	3.7849	Hex6dHex1HexNAc6
2393.849	2393.846	1.2532	Hex7HexNAc6
2421.770	2421.798	11.5617	Hex9HexNAc3NeuAc1 + 2Na
2487.867	2487.857	4.0195	Hex6dHex1HexNAc5NeuAc1 + 2Na

2539.902	2539.904	0.7874	Hex7dHex1HexNAc6
2685.955	2685.962	2.6061	Hex7dHex2HexNAc6
2743.062	2742.983	28.8008	Hex7dHex1HexNAc7
2759.059	2758.978	29.3587	Hex8HexNAc7
2905.054	2905.036	6.1961	Hex8dHex1HexNAc7
3124.155	3124.110	14.4041	Hex9HexNAc8

Table 5: N-Glycans Analyzed on the MALDI TOF. Included is the full m/z peak list of identified N-glycans on the MALDI TOF mass spectrometer.

As can be seen in Figure 10, many of the trends observed earlier regarding increased overall dysregulation, specifically of fucosylation, in S1 tumors are also seen in analysis of this cohort. For this analysis, data from S2/S3 tumors were combined, which was done to further examine whether S1 tumors were distinguishable from all other HCC tumors. S1 tumors exhibit increased expression of a larger number of unique glycan structures than S2/S3 tumors (Figure 10A). Structural trends regarding the association of increased branching/bisecting and the tumors remain for both S1 and S2/S3 tumors (Figure 10B). Tetra-antennary glycans make up a greater percentage of tumor-associated glycans in S2/S3 than S1, although some of higher mass fucosylated tetra-antennary structures that were detected on the FT-ICR instrument were not detected on the TOF instrument. Based on the tissues analyzed in this cohort and previously, it does not appear as though the correlation between branching and tumor progression is subtype dependent. However, fucosylated structures are overexpressed with more abundance and variety in S1 tumors than S2/S3 tumors, which validates observations regarding fucosylation in the discovery cohort (Figure 10C). Once again, it should be emphasized that these trends regarding fucosylation and glycan dysregulation are not universal within each subtype. There are S1 tumors in both cohorts that do not exhibit

the increased fucosylation being linked to S1, and there is a spectrum of S2/S3 tumors regarding the degree to which their glycosylation differs from surrounding tissue. This uncertainty can likely be attributed to both the imprecise nature of classifying all HCC cases into only three subtypes and the wide range of potential mechanistic effects through which altered glycosylation can support and drive tumor progression. However, trends regarding increased expression of branched structures in all three subtypes (Fig 10D) and increased expression of fucosylated structures in S1 (Fig 10E) do emerge and are worthy of further exploration for biomarker applications.

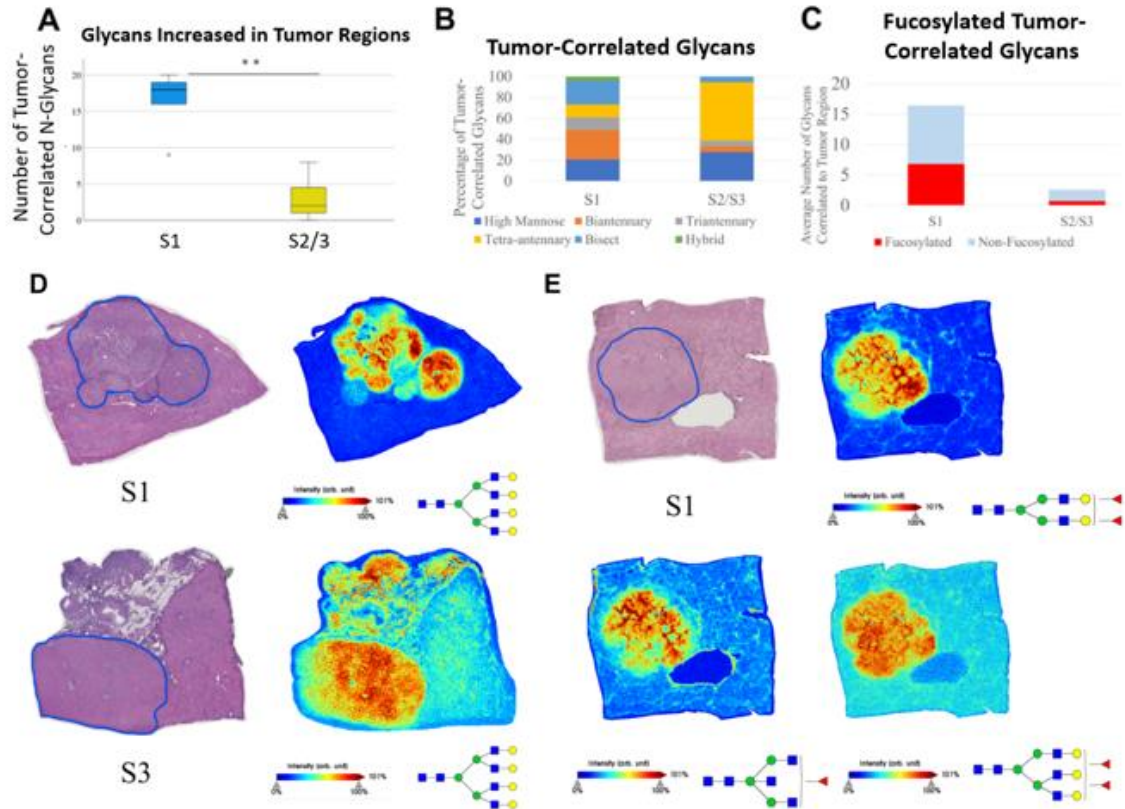


Figure 10. MALDI-TOF Validation Cohort. (A): Analysis of the number of unique glycan structures that are correlated (Pearson coefficient > 0.4) to the tumor region when compared to adjacent normal/cirrhotic region. n=5 (S1), n=2 (S2), n=5 (S3). Statistical test includes Wilcoxon Rank-Sum test (** = $p < 0.01$). (B): The structural features of glycans that are correlated to the tumor for S1 tumors and S2/S3 tumors (C): Analysis into the increase of fucosylated structures in S1 tumors and S2/S3 tumors. (D): Representative images of an S1 and S3 tumor to demonstrate that branched glycans without fucose are regularly overexpressed in tumors of all three subtypes, $m/z=2393.849$. The S3 tissue shown has a large region of fibrovascular tissue, which was not included in analysis but contains interesting glycosylation. (E): Representative images of an S1 tumor to demonstrate increased expression of fucosylated glycans, which extends to biantennary, bisecting, and branching structures. $m/z= 1955.699$ (top right), 1688.603 (bottom left), and 2320.835 (bottom right).

3.3.4 Discussion

Aberrant glycosylation has been considered a hallmark of cancer malignancy for many years.^{116,120,141,273,274} Mechanistically, glycosylation plays a critical role in cell signaling in non-cancerous tissue, and alterations to glycosylation motifs can drive abnormal signaling pathways in cancer, including increased growth factor signaling through increased core fucosylation as a prominent example.^{151,275} Additionally, expression of complex branched N-glycans has been linked to differentiation and metastasis along with the regulation of cytokine receptors.^{152,276,277} However, there is still a significant amount left unknown regarding both the nature of cancer-related glycosylation changes and the mechanism behind glycosylation's impact on cancer progression. Further exploration into these fields has promise to reveal both promising cancer markers and a new class of potential cancer drug targets.

This work focuses on the application of cancer-related N-glycosylation to the development of cancer markers, which is a rapidly developing field. Due to the ubiquitous nature of N-glycan expression on glycoproteins, the utilization of N-glycans as biomarkers for cancer could have profound implications on a wide range of cancer types. In many of these cases, molecular heterogeneity of cancer implies that the application of N-glycomics information to biomarker development is more likely to be successful as an algorithm of N-glycan expression as opposed to a single glycan structure, and there are already examples of this concept in development.⁸⁸ However, there is still a significant need to understand the inter-tumor heterogeneity of N-glycosylation in HCC in order to more specifically apply glycomics information, as current

glycomic biomarkers and algorithms have not been fully validated to be more sensitive for early stage HCC than AFP and/or ultrasound monitoring.⁶⁷

Previous work on HCC tissues has demonstrated that branching and fucosylation are increased on HCC tissue and serum, but there was significant glycan heterogeneity that made understanding how to apply this information difficult.^{40,139,245,278}

Incorporation of genetic and clinical information of each HCC tumor, which was novel to MALDI-IMS N-glycomic analysis, accounted for some of this heterogeneity. S1 tumors were demonstrated to have the most broadly dysregulated N-glycosylation in both cohorts analyzed, which allows this subtype to be distinguishable from other HCC tumors. Perhaps more significantly, fucosylated structures, which are conventionally considered to have increased expression overall in HCC, are increased with more frequency in S1 than in S2 or S3 tumors. This pattern of fucosylation in S1 tumors held up in both the cohort of samples analyzed through MALDI-FT-ICR and the cohort analyzed through the more clinically accessible MALDI-TOF. Interestingly, it appeared as though fucosylation may actually be decreased in S2 tumors, based upon both a lack of fucosylated structures with increased tumor expression and increased expression of fucosylated structures outside the tumor, which is in stark contrast to what is seen in S1 tumors considering their similarity in clinical outcomes. This suggests that aggressive stromal tumors have increased fucosylation within the tumor, whereas aggressive stemness tumors have decreased fucosylation within the tumor.

Drastic overexpression of a multitude of varied glycan structures, many including fucose(s), appears to be limited to S1 tumors, as only one S2/S3 tumor had more than

five glycans with a threefold increase in the tumor, and only one S2/S3 tumor had more than three glycans determined to be spatially correlated to the tumor region. However, aberrant glycan expression is not uniform in S1 tumors, which is demonstrated by two S1 tissues, one in each cohort, that exhibited an increase to primarily non-fucosylated glycans. Further analysis into the specific features of S1 tumors that correlate with increased expression of complex fucosylated structures will be necessary. However, this data suggests that S1 tumors encapsulate most, if not all, of the HCC tumors that demonstrate extremely different glycosylation from adjacent cirrhotic liver tissue, specifically including increased fucosylation. Further validating this finding could prove very useful towards applying glycosylation information clinically as a method of using fucosylation patterns to distinguish S1 tumors easily.

Both S2 and S3 tumors demonstrate N-linked glycosylation that is more consistent with that of the adjacent, non-HCC tissue. There are typically still glycan structures in each tumor that exhibit increased expression, and these are often branched structures which have previously been shown to be directly linked to cancer progression in a variety of cancer types.^{147,152,276} S3 tumors having similar glycosylation to their surrounding tissue is unsurprising, as this subclass of tumors retains a hepatocyte phenotype, is well-differentiated, and includes the least aggressive HCC tumors. S3 tumors also have fewer hallmark aberrant signaling pathways, which altered glycosylation would play important roles in. However, these clinically favorable features of S3 tumors are not the case in S2 tumors, which are poorly differentiated and aggressive in nature. Along with poor clinical outcomes, S1 and S2 tumors share several

mutated signaling pathways, including through canonical WNT signaling and mutated p53 signaling. Based on this N-glycosylation data, S2 tumors progress into rapidly proliferating, metastatic tumors in a very different manner than S1 tumors, specifically in a manner that does not rely on the overexpression of many fucosylated N-glycan structures. Further work into mechanistic differences between these subtypes could be exceedingly valuable in gaining an understanding into the relation of N-glycosylation and HCC development, progression, and metastasis, and it could offer more specific targets for both biomarker and drug development.

The validation of increased fucosylation in S1 tumors would have profound impacts for biomarker development. One of the features used to classify S2 tumors is significantly increased AFP levels from the other two subtypes, and AFP has already been demonstrated to have clinical biomarker utility.¹⁶⁸ If AFP could be supplemented with a fucosylated marker that could increase detection of S1 tumors, which have lower AFP levels, this would have a notable impact on HCC screening effectiveness and HCC mortality reduction. Indeed, our work has identified fucosylated kininogen as a good partner for AFP in biomarker algorithms and the results presented here may explain why this is the case.^{89,184,185,244} S3 tumors having less dramatic glycosylation abnormalities was expected, as this subtype is well-differentiated and less aggressive. Therefore, these tumors may be more indolent in nature with slower tumor doubling times and potentially lower risk of HCC-related mortality.²⁷⁹ Combining glycomic differences of S1 tumors and the high AFP levels of S2 tumors could lead to better detection of the most aggressive tumors, which would have profound effects on overall survival of HCC

patients. In order for this to become clinically valuable, the distinguishing N-glycosylation features discussed here must be validated in serum.

Overall, there is clear value in incorporating genetic and clinical information into analysis of N-glycosylation of HCC, and it would likely be worthwhile to expand this idea to other cancer types. Biomarker development for HCC as a whole has largely lagged behind that of other cancer types, and there is still heavy reliance on imaging techniques for surveillance and detection. In order to develop specific markers of HCC, there must be incorporation of more specific information on each tumor beyond simply survival data. HCC is incredibly diverse molecularly and genetically, which suggests that successful markers for HCC must be able to detect a diverse range of presentations of HCC. Discovering markers for specific subclasses of HCC makes it more likely that these markers will complement each other to increase sensitivity to early-stage HCC as a whole, which is the most promising avenue to reduce HCC-related mortality.

Chapter 4: Novel Enzymatic Approach to Analyze De- Sialylated N-linked Glycans

This chapter has been adapted from a manuscript published in the *Journal of Proteome Research*, 2022 21 (8): 1930-1938. ATD performed experiments, data analysis, and writing of the manuscript. HL performed enzyme production/purification. PA, RD, and AM contributed intellectually to the manuscript.

**Novel combined enzymatic approach to analyze non-sialylated N-linked glycans
through MALDI Imaging Mass Spectrometry**

Andrew T DelaCourt, Hongyan Liang, Richard R. Drake, Peggi M. Angel, Anand S. Mehta

4.1 Introduction

Glycosylation of glycoproteins, specifically N-linked glycosylation, has been well established to have important functional roles regarding cell growth and signaling, and aberrant glycosylation has been directly linked to the progression of a variety of disease states, including cancer.^{44,120,139,153,242,261,280–283} Analysis of N-glycan expression has historically focused more on serum glycoproteins due to ease of sample collection and relative abundance for glycan structural analysis, but recently developed matrix-assisted laser desorption ionization imaging mass spectrometry (MALDI-IMS) techniques allow for glycomic analysis of tissue samples in addition to serum samples in a high-throughput manner.^{234,235,270,278,284–286} This MALDI-IMS technique allows for comprehensive glycomic analysis of enzymatically released N-glycan structures on both tissue and serum samples.^{245,287,288} Particularly regarding the analysis of tissue samples, retaining spatial localization is critical to analyzing how glycan structures colocalize with distinct histological tissue regions and cellular features.¹¹⁵ This methodology was originally developed utilizing the enzyme peptide-N-glycosidase F Prime™ (PNGase F) to cleave N-glycans indiscriminately from associated glycoproteins.²¹⁹ Recently, the methodology has been adapted for alternative endoglycosidases in order to add structural glycomic information and to multiplex enzymes to analyze both glycosylation and extracellular matrix (ECM) content on the same tissue sample.^{289,290} Utilizing additional enzymes to PNGase F can be done with sequential enzymatic digestions, such as endoglycosidase F3 (Endo F3) Prime™ to target core fucosylated N-glycans, or with

concurrent enzymatic digestions to improve sensitivity for specific structural features of N-glycan analytes via MALDI-IMS.

Sialidases, also termed neuraminidases, are a family of exoglycosidases that cleave nonreducing sialic acid residues from their associated glycoconjugates, including N-glycans.^{291,292} Sialidases are expressed in mammalian, viral, and bacterial species, and varying sialidases have differing catalytic efficiency based on the glycoconjugate and the glycosidic linkage.²⁹³ Sialidases that target N-glycans cleave terminal sialic acid residues from their associated glycoconjugates while leaving the remainder of the glycan structure intact. Sialylation of glycoproteins is well-established to have important roles in cellular recognition, cell adhesion, and cell signaling, and aberrant sialylation has been observed in several tumor types.^{116,150} However, there are cancers in which the dominant N-glycan changes do not include sialylation, but instead fucosylation and/or branching of glycan structures.^{115,138,245} Fucosylation in particular has been demonstrated to be altered in both serum and tissue of a number of cancer types including breast, liver, ovarian, prostate, colorectal and lung, and these fucose-specific altered expression patterns are often independent of sialylation.¹¹⁹

With the utilization of MALDI-IMS to analyze N-glycan structures, both sialylated and non-sialylated N-glycan species can be readily observed. While the stability of sialic acids can be an issue in certain sources, this is less of a problem in newer MALDI FT-ICR instruments with softer ionization conditions and a cooling gas.^{127,278,294,295} Detecting these N-glycans with enough sensitivity to analyze their spatial distribution is critical to the analysis of their role in disease presence/progression. However, there are some

lower abundant N-glycan structures that are oftentimes not observed with adequate sensitivity through MALDI-IMS to confidently analyze their spatial localization within a tissue, which is particularly true of higher molecular weight, complex glycans. The presence of sialic acid residues on these large, complex N-glycan structures substantially increases the molecular weight of the glycan, which in some cases pushes the glycan outside of the mass detection range of established MALDI-IMS techniques.

Biosynthesized N-glycans generally range from 1000-5000 m/z , although the readily observed N-glycans through the established MALDI-IMS methodology on tissue and in serum are typically less than 3500 m/z .^{94,219,234,278} With this in mind, we adapted a MALDI-IMS N-glycan workflow to incubate concurrently with sialidase and PNGase F, with aims to increase sensitivity for non-sialylated N-glycan structures and identify key N-glycan structural scaffolds that are frequently sialylated. This simultaneous enzyme treatment will result in m/z peaks for sialylated N-glycan structures being shifted (through the loss of 291 Da plus a sodium adduct) to the m/z of a corresponding, non-sialylated N-glycan. This alone will result in increased signal intensity for detected N-glycans, and sensitivity can increase even further with improved enzyme efficiency of endoglycosidases on-tissue due to the removal of sialic acids from the tissue surface. This workflow will be especially applicable to studies targeting fucosylation or the increased expression of large, branched N-glycans. Therefore, the proposed workflow is expected to significantly increase MALDI-IMS sensitivity for non-sialylated N-glycans, and it is highly relevant for the analysis of N-glycosylation independent of sialylation status.

4.2 Materials and Methods

4.2.1 Cloning, Expression, and Purification of a Recombinant Sialidase

The cDNA fragment covering the first 1485bp (495 amino acids) of the coding sequence (which is 990 amino acids) of *Arthrobacter ureafaciens* sialidase (EC 3.2.1.18) was generated by PCR and 10xHis was added to its C terminal.²⁹⁶ Amplified DNA fragments were cloned into pQE-60 by NcoI/XbaI (Genscript, Piscataway, NJ). The constructed plasmid, pQE-60 SA-10xHis, was transformed into BL21 (DE3). The transformants were cultured in LB broth supplemented with 100 µg/ml Ampicillin. Cultures were grown at 37 °C until the cells reached an A600 nm of about 0.5, 0.5 mM IPTG were added to the culture to induce protein overproduction at 20 °C. After overnight incubation, the cells were harvested by centrifugation. The cell pellets were re-suspended using PBS plus Pierce protease inhibitor tablets (Thermo Fisher Scientific, Waltham, MA), stored at -20 °C. Omnicleave endonuclease (Lucigen Corporation, Middleton, WI) and MgCl₂ were added to thawed cell suspension. The cell suspension was incubated at room temperature for at least one hour with rocking. The cells were lysed using French Press (GlenMills Inc., Clifton, NJ) per the manufacturer's instructions. The cell lysate was applied to HisTrap FF (GE Healthcare, Pittsburgh, PA) and washed with 20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole (pH 7.4). Bound His-tagged protein was eluted with a gradient from 120 to 420 mM imidazole in 20 mM sodium phosphate, 0.5 M NaCl (pH 7.4). The purified sialidase was desalted and concentrated with 20 mM Tris-HCl, 50 mM NaCl (pH 7.5) using Spin-X UF Concentrator (10kDa; Corning). The protein purity was confirmed using SDS-PAGE to be >95%.

4.2.2 In-Solution Digestion by Recombinant Sialidase

Bovine Fetuin (New England BioLabs) was incubated with recombinant sialidase, now Sialidase Prime™ from N-Zyme Scientifics (Doylestown, PA), or α 2-3,6,8 Neuraminidase (New England BioLabs) at an enzyme-to-protein-ratio of 1:5 (w/w) at pH 5.5, 37 °C for 1 hour.

4.2.3 N-Glycan Sequencing

Human Fetuin-A (Assaypro, St. Charles, MO) was run on SDS-PAGE gel (SimplyBlue SafeStain, from Invitrogen), stained, and cut out. 5 μ L of normal human serum (Sigma) were absorbed into gel plugs. The fetuin-A gel pieces or serum gel plugs were alkylated in the dark for 30 min with iodoacetamide, fixed in a solution of 10% methanol 7% acetic acid for 1 hour, washed in acetonitrile, followed by subsequent steps of 20 mM ammonium bicarbonate (pH 7.0) and acetonitrile before being dried in a speed-vac. PNGase F or Endo F3 was diluted with corresponding buffer and allowed to absorb into and cover the gel pieces, incubate overnight at 37 °C. The glycans were eluted from the gel pieces by sonication in Milli-Q water, dried down and labeled with a 2AB dye. The glycans were cleaned up using paper chromatography and filtered using PTFE syringe filter unit. Fluorescently labeled glycans were separated on normal phase Waters Alliance UPLC system as previously described.²⁹⁷ Samples were further digested with Sialidase Prime for calculation of glucose unit (GU) value and compared to GlycoStore database (www.glycostore.org).

4.2.4 Patient Tissues

Formalin fixed paraffin embedded (FFPE) tissue cores (5 μ M thickness) were made from 4 hepatocellular carcinoma (HCC) tissue blocks from surgical resection in the background of liver cirrhosis. Samples were provided from the UT Southwestern Medical Center under IRB 102010-051 to Dr. Amit Singal. The glycan imaging work at MUSC was performed under IRB Pro00079936. All tissues were H&E stained following MALDI-IMS analysis.

4.2.5 FFPE Tissue Preparation for MALDI-IMS

UPLC grade methanol (Catalog No. A452SK-4), ethanol (Catalog No. 22-032-601), acetonitrile (Catalog No. A998-1), xylene (Catalog No. X3S-4), and water (Catalog No. W5-1) were obtained from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid (Catalog No. W302031) and α -cyano-4- hydroxycinnamic acid (Catalog No. C89892) were obtained from Sigma-Aldrich (St. Louis, MO). Peptide-N-glycosidase F (PNGase F) PrimeTM and Endoglycosidase F3 (Endo F3) PrimeTM were cloned, expressed, and purified from N-Zyme Scientifics (Doylestown, PA) as previously described.²¹⁹

FFPE tissues were prepared according to a previously published protocol, with added incubation of Sialidase Prime concurrently with PNGase F Prime incubation.^{219,234,270} Tissue Tack microscope slides were purchased from Polysciences Inc (Warrington, PA, Catalog No. 24216), and indium tin oxide glass slides were purchased from Delta Technologies (Loveland, CO, Catalog No. CB-40IN-S111). Briefly, slides were heated to 60°C for 1 hour, washed with xylenes, and rehydrated with a series of ethanol

and water washes. Slides were then processed by antigen retrieval, heating to 95° C for 30 minutes in a Decloaking Chamber in a 10 mM citraconic anhydride buffer, pH 3. Slides were cooled to room temperature and buffer exchange was done to replace buffer with 100% water. Enzymes (PNGase F Prime, Endo F3 Prime, and Sialidase Prime) were applied at a concentration of 0.1 mg/mL each in HPLC grade water using a M5 TM-Sprayer Tissue MALDI Sample Preparation System (HTX Technologies, LLC). After a two-hour enzyme incubation at 37 °C, MALDI matrix α -cyano-4-hydroxycinnamic acid (0.042 g CHCA in 6 mL 50% acetonitrile/49.9% water/0.1% TFA) was sprayed by the M5 TM-Sprayer.

4.2.6 N-Glycan Imaging via MALDI-IMS

As previously described, N-glycans were analyzed using a MALDI FTICR mass spectrometer (SolariX Dual Source, 7T, Bruker Daltonics, m/z 500-5000).²¹⁹ The data was then analyzed and visualized using FlexImaging 5.0 and SCiLS Lab 2022a (Bruker Daltonics). Putative N-glycan structures are shown based on previous databases built with use of GlycoWorkBench (RRID:SCR_000782).²⁷¹ Glycan structures along with associated m/z values and mass error can be found in Tables 6 and 7.

Observed m/z	Theoretical m/z	Error in PPM	Glycan Structure
933.3145	933.317	2.6786	Hex3HexNAc2
1079.3769	1079.3749	1.8529	Hex3dHex1HexNAc2
1095.3604	1095.3698	8.5816	Hex4HexNAc2
1136.3965	1136.3964	0.0880	Hex3HexNAc3
1257.4237	1257.4226	0.8748	Hex5HexNAc2
1282.4572	1282.4543	2.2613	Hex3dHex1HexNAc3
1298.4492	1298.4492	0.0000	Hex4HexNAc3
1339.4734	1339.4757	1.7171	Hex3HexNAc4
1419.4743	1419.4755	0.8454	Hex6HexNAc2
1444.5122	1444.5071	3.5306	Hex4dHex1HexNAc3
1460.4982	1460.502	2.6018	Hex5HexNAc3
1485.5325	1485.5337	0.8078	Hex3dHex1HexNAc4
1501.5259	1501.5286	1.7982	Hex4HexNAc4
1542.56	1542.5551	3.1765	Hex3HexNAc5
1581.5245	1581.5282	2.3395	Hex7HexNAc2
1606.5486	1606.5599	7.0337	Hex5dHex1HexNAc3
1611.5159	1611.5266	6.6397	Hex4HexNAc3NeuAc1+ 2Na
1622.5427	1622.5548	7.4574	Hex6HexNAc3
1647.5832	1647.5865	2.0029	Hex4dHex1HexNAc4
1663.5855	1663.5814	2.4646	Hex5HexNAc4
1688.5998	1688.613	7.8171	Hex3dHex1HexNAc5
1704.5994	1704.608	5.0451	Hex4HexNAc5
1743.5798	1743.5810	0.6882	Hex8HexNAc2
1809.6378	1809.6393	0.8289	Hex5dHex1HexNAc4
1850.6595	1850.6659	3.4582	Hex4dHex1HexNAc5
1866.6597	1866.6608	0.5893	Hex5HexNAc5
1905.6406	1905.6338	3.5684	Hex9HexNAc2
1954.6758	1954.6768	0.5116	Hex5HexNAc4NeuAc1
1955.697	1955.6972	0.1023	Hex5dHex2HexNAc4
1976.6662	1976.6666	0.2024	Hex5HexNAc4NeuAc1 + 2Na
2012.7238	2012.7187	2.5339	Hex5dHex1HexNAc5
2028.7139	2028.7136	0.1479	Hex6HexNAc5
2069.7236	2069.7401	7.9720	Hex5HexNAc6
2100.7302	2100.7347	2.1421	Hex5dHex1HexNAc4NeuAc1
2101.7347	2101.7551	9.7062	Hex5dHex3HexNAc4
2122.7303	2122.7245	2.7323	Hex5dHex1HexNAc4NeuAc1 + 2Na
2157.7399	2157.7562	7.5541	Hex5HexNAc5NeuAc1
2158.7941	2158.7766	8.1064	Hex5dHex2HexNAc5
2174.7771	2174.7715	2.5750	Hex6dHex1HexNAc5
2245.7677	2245.7722	2.0038	Hex5HexNAc4NeuAc2
2246.7816	2246.7926	4.8959	Hex5dHex2HexNAc4NeuAc1
2267.7743	2267.7542	8.8634	Hex5HexNAc4NeuAc2 + 2Na
2268.787	2268.7746	5.4655	Hex5dHex2HexNAc4NeuAc1 + 2Na
2289.7338	2289.7361	1.0045	Hex5HexNAc4NeuAc2 + 3Na
2303.8054	2303.8141	3.7763	Hex5dHex1HexNAc5NeuAc1

2304.8153	2304.8345	8.3303	Hex5dHex3HexNAc5
2319.7991	2319.809	4.2676	Hex6HexNAc5NeuAc1
2320.8493	2320.8294	8.5745	Hex6dHex2HexNAc5
2325.8052	2325.7961	3.9126	Hex5dHex1HexNAc5NeuAc1 + 2Na
2341.7867	2341.7988	5.1670	Hex6HexNAc5NeuAc1 + 2Na
2391.8158	2391.8301	5.9787	Hex5dHex1HexNAc4NeuAc2
2393.841	2393.8458	2.0051	Hex7HexNAc6
2413.799	2413.8121	5.4271	Hex5dHex1HexNAc4NeuAc2 + 2Na
2421.7472	2421.7478	0.2478	Hex9HexNAc3NeuAc1 + 2Na
2435.8005	2435.794	2.6685	Hex5dHex1HexNAc4NeuAc2 + 3Na
2465.8524	2465.8669	5.8803	Hex6dHex1HexNAc5NeuAc1
2487.8388	2487.8489	4.0597	Hex6dHex1HexNAc5NeuAc1 + 2Na
2539.9092	2539.9037	2.1654	Hex7dHex1HexNAc6
2611.937	2611.9248	4.6709	Hex6dHex2HexNAc5NeuAc1
2632.8924	2632.8864	2.2789	Hex6HexNAc5NeuAc2 + 2Na
2633.8954	2633.9068	4.3282	Hex6dHex2HexNAc5NeuAc1 + 2Na
2684.9217	2684.9412	7.2627	Hex7HexNAc6NeuAc1
2685.9385	2685.9616	8.6003	Hex7dHex2HexNAc6
2852.9591	2852.9811	7.7112	Hex7dHex1HexNAc6NeuAc1 + 2Na
2905.0249	2905.0359	3.7865	Hex8dHex1HexNAc7

Table 6. N-Glycan Peak List with PNGase F. Full list of glycans analyzed on the FT-ICR mass spectrometer with PNGase F. All analytes are sodiated.

Observed <i>m/z</i>	Theoretical <i>m/z</i>	Error in PPM	Glycan Structure
933.3124	933.317	4.9287	Hex3dHex1HexNAc3
1095.3636	1095.3698	5.6602	Hex4dHex1HexNAc3
1136.3908	1136.3963	4.8399	Hex3dHex1HexNAc4
1257.421	1257.4226	1.2724	Hex5dHex1HexNAc3
1298.4512	1298.4492	1.5403	Hex4dHex1HexNAc4
1339.4806	1339.4757	3.6581	Hex3dHex1HexNAc5
1419.473	1419.4754	1.6908	Hex6dHex1HexNAc3
1444.5058	1444.5071	0.9000	Hex4dHex2HexNAc4
1460.5008	1460.502	0.8216	Hex5dHex1HexNAc4
1501.5296	1501.5286	0.6660	Hex4dHex1HexNAc5
1581.5235	1581.5283	3.0350	Hex7dHex1HexNAc3
1589.5558	1589.5446	7.0460	Hex4dHex1HexNAc4NeuAc1
1606.5613	1606.5599	0.8714	Hex5dHex2HexNAc4
1611.5176	1611.5265	5.5227	Hex4dHex1HexNAc4NeuAc1 + 2Na
1663.5871	1663.5814	3.4263	Hex5dHex1HexNAc5
1743.5829	1743.5811	1.0324	Hex8dHex1HexNAc3
1751.5905	1751.5974	3.9393	Hex5dHex1HexNAc4NeuAc1
1773.5872	1773.5794	4.3979	Hex5dHex1HexNAc4NeuAc1 + 2Na
1809.6533	1809.6393	7.7363	Hex5dHex2HexNAc5
1825.6413	1825.6342	3.8891	Hex6dHex1HexNAc5
1919.6309	1919.6373	3.3340	Hex5dHex2HexNAc4NeuAc1 + 2Na
1971.707	1971.6921	7.5570	Hex6dHex2HexNAc5
1976.6734	1976.6588	7.3862	Hex5dHex1HexNAc5NeuAc1 + 2Na
2028.7319	2028.7136	9.0205	Hex6dHex1HexNAc6
2064.6694	2064.6748	2.6154	Hex5dHex1HexNAc4NeuAc2 + 2Na
2086.6578	2086.6567	0.5272	Hex5dHex1HexNAc4NeuAc2 + 3Na
2138.7278	2138.7116	7.5747	Hex6dHex1HexNAc5NeuAc1 + 2Na
2190.7756	2190.7664	4.1994	Hex7dHex1HexNAc6
2284.7847	2284.7695	6.6527	Hex6dHex2HexNAc5NeuAc1 + 2Na
2451.8066	2451.789	7.1784	Hex6dHex1HexNAc5NeuAc2 + 3Na
2503.8645	2503.8438	8.2673	Hex7dHex1HexNAc6NeuAc1 + 2Na

Table 7. N-Glycan Peak List with Endo F3. Full list of glycans analyzed on the FT-ICR mass spectrometer with Endo F3. All analytes are sodiated.

4.2.7 UPLC of Tissue Extracted N-Glycans

After sialidase/PNGase F enzyme application on HCC liver tissues, slides were incubated for 2 hours at 37 °C to complete enzymatic digestion. Following incubation, 100 µL of HPLC grade water was applied on top of the tissue and incubated for 20 minutes to extract enzymatically released N-glycans. The water was removed from the tissue and concentrated under vacuum by centrifugation. Concentrated glycans were labeled with a 2AB dye. The glycans were cleaned up using paper chromatography and filtered using PTFE syringe filter unit. Fluorescently labeled glycans were separated on normal phase Waters Alliance UPLC system as previously described.²⁹⁷ GU units were calculated and compared to GlycoStore database (www.glycostore.org).

4.3 Results

4.3.1 In-Solution Analysis of Sialidase Prime™ Activity on N-Glycans

The activity of purified recombinant sialidase (Sialidase Prime™), from the bacterium *Arthrobacter ureafaciens*, was initially validated through a comparison to the activity of a commercially available α 2-3,6,8 Neuraminidase, from *Clostridium perfringens*, as shown in Figure 11. This validation was done by comparing enzymatic activity of both enzymes targeting bovine fetuin. Fetuin is a glycoprotein containing sialylated N-glycan and O-glycan structures that is commonly used as a positive control for glycosidase enzymes, and sialylation is well-established to be important to fetuin activity.²⁹⁸ As shown by SDS-PAGE, Sialidase Prime will cleave sialic acids from fetuin with similar activity as commercially available neuraminidase, which is known to efficiently cleave sialic acids from a variety of glycoproteins and glycosidic linkages. The lack of additional band shifts caused by Sialidase Prime compared to the commercially available neuraminidase supports the claim that Sialidase Prime does not have off target glycosidase activity beyond targeting sialic acids. As validated by SDS-PAGE, Sialidase Prime and the commercially available neuraminadase enzyme are differing molecular weights, with Sialidase Prime expected to be 54 kDa and the neuraminadase at 43 kDa, which is less than the full-length proteins.

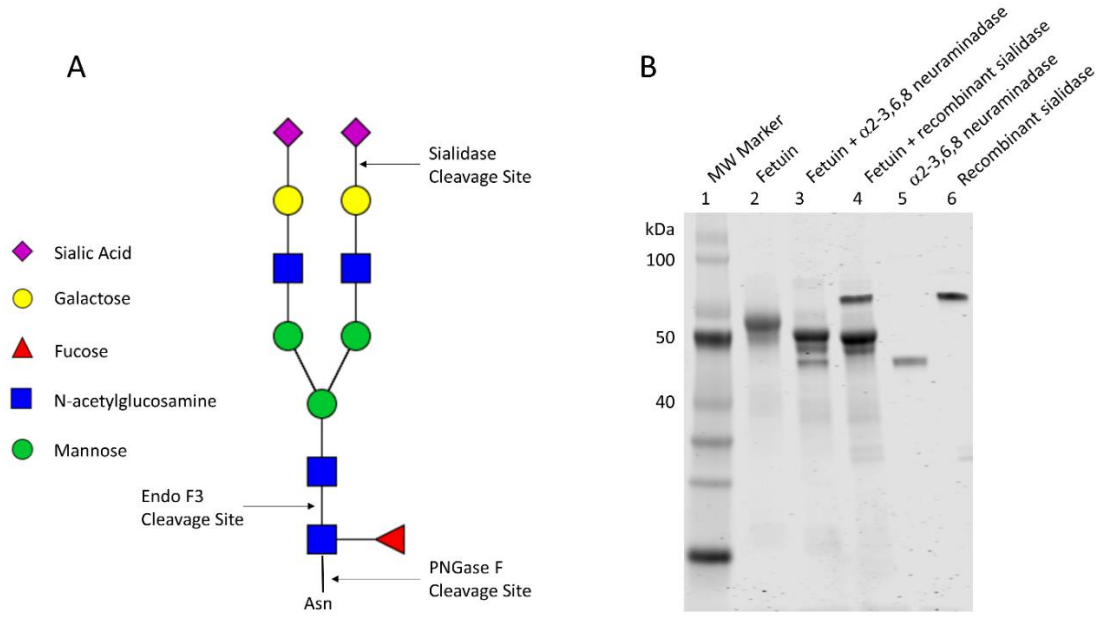


Figure 11. Sialidase Prime Activity Validation. (A): A schematic representation of N-glycan structures, with cleavage sites of the three enzymes utilized in this work. (B): SDS-PAGE of our recombinant sialidase and α 2-3,6,8 neuraminidase on bovine fetuin. There are comparable band shifts between the two enzymes, demonstrating sialic acid removing activity of Sialidase Prime.

The enzymatic activity of Sialidase Prime was also investigated by normal-phase UPLC, as shown in Figure 12. N-glycans were enzymatically released using either PNGase F to target all N-glycans or Endo F3 to target core fucosylated N-glycans, and simultaneous incubation with Sialidase Prime led to the removal of terminal sialic acids from the released glycan structures. Peaks were identified utilizing a glucose homopolymer standard to calculate glucose units (GU), which were then compared to a published glycan database. It is important to note that for the same glycan identification, retention time/glucose units differ between Endo F3 cleaved glycans and PNGase F cleaved glycans due to differing cleavage sites. UPLC analysis of human fetuin A demonstrated the common glycan structure A2G2 and both A2G2S1 and A2G2S2 when treated with only PNGase F or Endo F3 (Figure 12A). Additional treatment with sialidase led to the removal of A2G2S1 and A2G2S2 and a corresponding increase in intensity to A2G2. The same experiment utilizing normal, pooled human serum led to similar results, with sialidase treatment removing sialylated peaks that were present in the chromatogram with only PNGase F/Endo F3 treatment (Figure 12B). This demonstrates that Sialidase Prime can effectively cleave sialic acids from enzymatically released N-glycan structures originating from serum glycoproteins, which causes an increase in the measurement of non-sialylated N-glycan structures. Overall, this demonstrates that Sialidase Prime effectively removes sialic acids of N-glycan structures in combination with PNGase F/Endo F3 N-glycan cleaving activity.

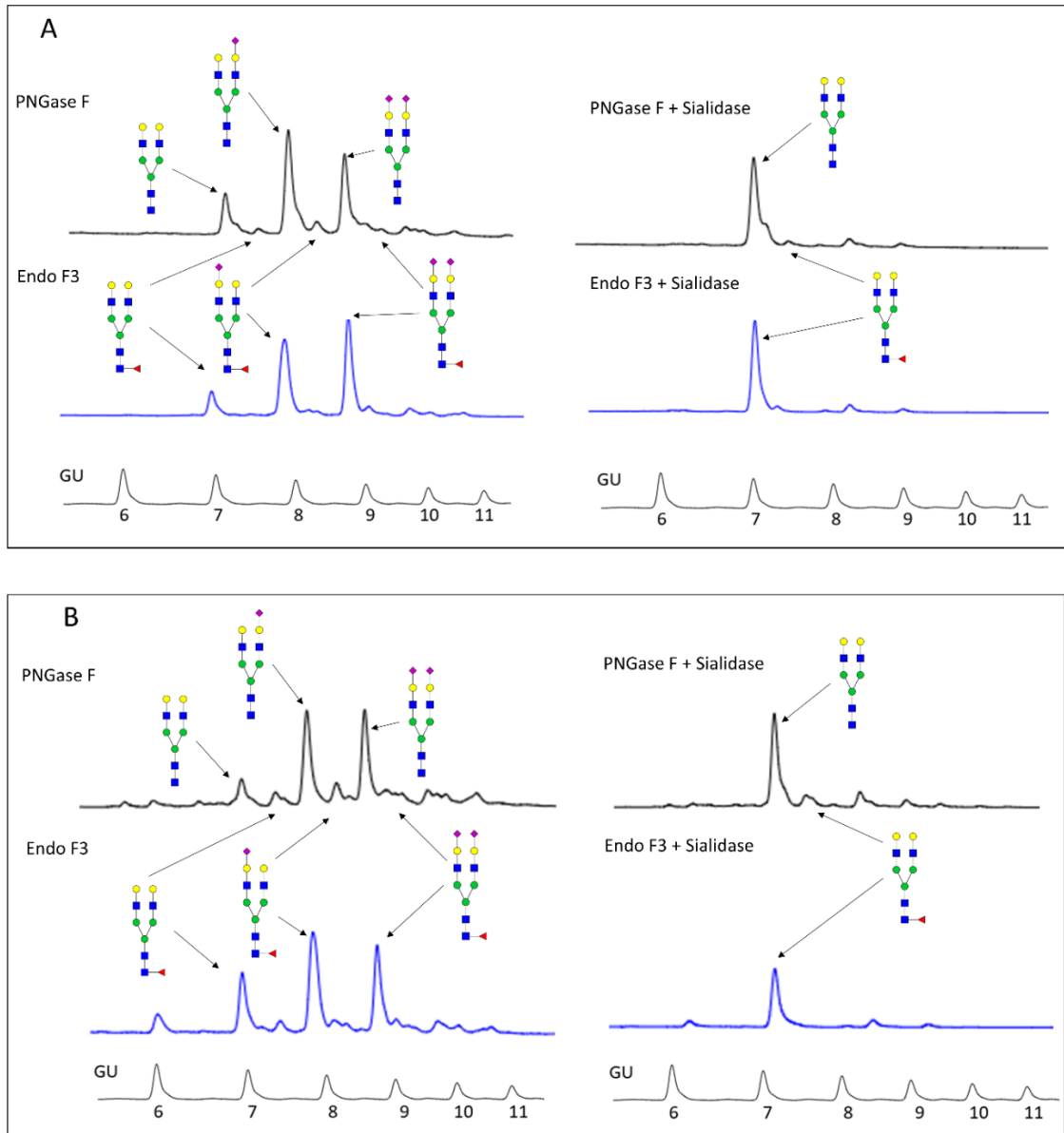


Figure 12. Sialidase Prime Activity on Cleaved N-Glycans. (A): PNGase F and Endo F3 activity with and without simultaneous sialidase incubation targeting human fetuin A. (B): PNGase F and Endo F3 activity with and without simultaneous sialidase incubation targeting pooled, normal human serum.

4.3.2 On-Tissue Analysis of Sialidase Activity using MALDI-IMS

After validating the activity of Sialidase Prime in solution, we then utilized the enzyme within our previously developed MALDI-IMS N-glycan methodology.^{219,290} The workflow through which tissues slides were prepared for MALDI-IMS is identical to previously described MALDI-IMS N-glycan imaging (Figure 5, Chapter 3), with the addition of a concurrent Sialidase Prime incubation to the PNGase F incubation. H&E stains of serial sections of the tissues analyzed are included in Figure 13. The on-tissue MALDI-IMS methodology was first utilized on two unique HCC tissues (three serial sections each) to examine PNGase F Prime only treatment, Sialidase Prime only treatment, and combined enzymatic treatment (Figure 14). Sialidase only enzymatic treatment served as a negative control, as full N-glycan structures were not enzymatically released from their associated glycoproteins. The m/z peak of the most abundant sialylated N-glycans, as determined by PNGase F only treatment, was nearly completely eliminated with combined PNGase F/sialidase treatment (Figure 14A-C). This suggests that combined PNGase F/sialidase treatment is sufficient to remove sialic acids from observed N-glycans, and thereby eliminate m/z peaks of sialylated N-glycan structures. Meanwhile, the most abundant non-sialylated N-glycan of the tissue was substantially increased with combined PNGase F/sialidase treatment compared to PNGase F treatment alone (Figure 14D-F). It is important to note that these m/z images were normalized to total ion count (TIC) together, and the spatial distribution of tissues with lower signal intensity (PNGase F only treatment) can be better observed in m/z images that were normalized separately. Spectra demonstrating eliminated sialylated N-

glycan expression with sialidase and increased N-glycan expression of abundant non-sialylated N-glycans are shown in Figure 14G-H. When analyzing all detected N-glycans, both tissues analyzed in this experiment demonstrate increased area under the peak (AUP) of nearly all non-sialylated glycan structures for combined enzyme treatment, including N-glycans that would not be expected to be sialylated, and there was a nearly complete reduction in AUP of sialylated N-glycan structures down to the level of background signal observed in sialidase-only treated tissue (Table 8). In the majority of cases, the percent increase in AUP of non-sialylated structures was substantially more than the observed AUP of structurally associated sialylated N-glycans, suggesting that sialidase treatment leads to increased signal intensity overall as opposed to simply combining associated sialylated and non-sialylated m/z peaks.

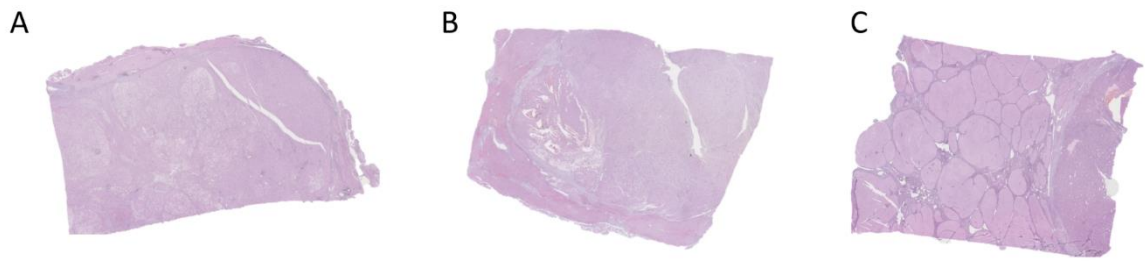


Figure 13. Tissue H&E Stains. (A/B): Tissues included in Figure 14 sprayed with PNGase F/Sialidase. (C): Tissue included in Figure 15 sprayed with Endo F3/Sialidase and PNGase F.

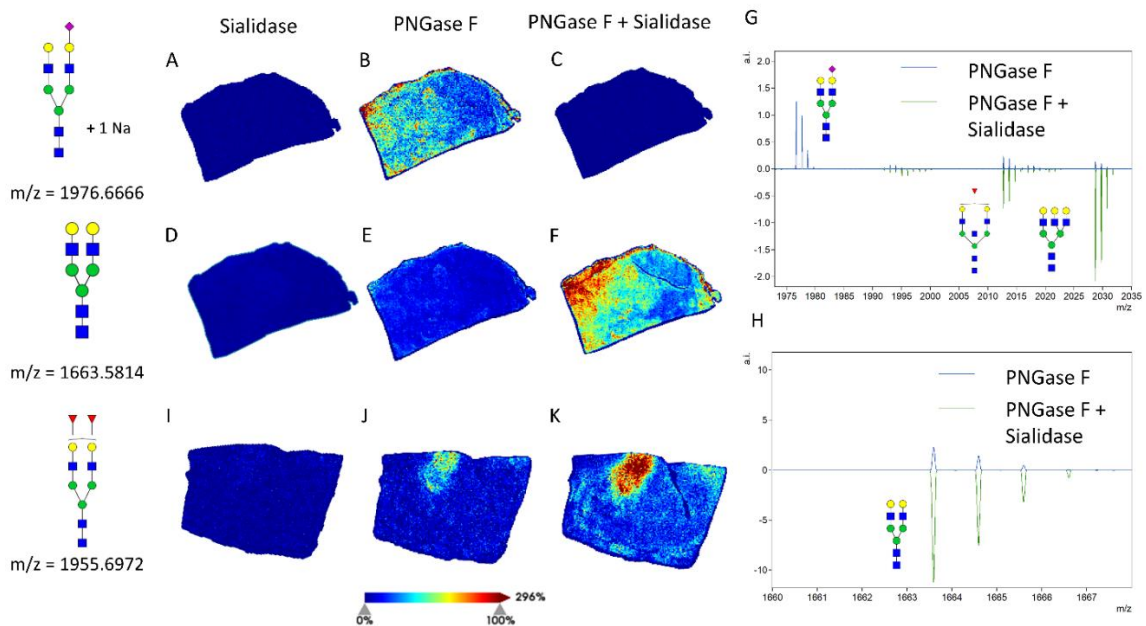


Figure 14. MALDI-IMS with Simultaneous PNGase F and Sialidase Incubation. (A-C): m/z images from an HCC tissue of the most highly abundant sialylated N-glycan with enzyme treatments including sialidase only, PNGase F only, and a combination. (D-F): m/z images demonstrating the increase in signal intensity with combined enzyme treatment of the most highly abundant N-glycan on the tissue. (G-H): Mass spectra with both PNGase F treatment and combined PNGase F/Sialidase treatment are shown to demonstrate the increased signal intensity of non-sialylated N-glycan peaks and the decreased signal intensity of sialylated N-glycan peaks. (I-K): m/z images of a multi-fucosylated N-glycan structure, which would be interfered with by the +1 isotope of A2G2S1 without sialidase activity to remove the potential interfering peak, is shown on a second HCC tissue.













Observed <i>m/z</i>	Proposed Glycan Structure	Composition	AUP PNGase F Tissue 1	AUP PNGase F/Sialidase Tissue 1	Log2 Fold Change Tissue 1	AUP PNGase F Tissue 2	AUP PNGase F/Sialidase Tissue 2	Log2 Fold Change Tissue 2
1136.4092		A1	37217	82375	1.15	97088	126571	0.38
1663.5772		A2G2	1026966	4140239	2.01	3727112	6936026	0.90
1809.6416		FA2G2	747880	1879706	1.33	2192912	4769753	1.12
1866.6645		A2BG2	48213	105752	1.13	143334	346675	1.27
1976.6647		A2G2S1	681608	19387	-5.14	22098	17866	-0.31
2012.7544		FA2BG2	213018	423978	0.99	560140	1636322	1.55
2028.7689		A3G3	186381	958632	2.36	1080310	2215655	1.04
2174.7872		FA3G3	226977	857240	1.92	1559926	4065374	1.38
2391.8121		FA2G2S2	54465	36040	-0.60	147308	39276	-1.91
2393.8178		A4G4	70800	347018	2.29	277180	657013	1.25
2539.8411		FA4G4	81951	412425	2.33	652820	1816432	1.48
2904.9332		F(6)A4G(4)4Lac1	91639	144805	0.66	242627	557655	1.20

Table 8. Signal Intensity with PNGase F and with PNGase F/Sialidase. Included are selected N-glycans with notable signal intensity differences between PNGase F only treatment and PNGase F/sialidase combined enzymatic treatment. This includes sialylated N-glycans with decreased signal intensity and non-sialylated N-glycans with increased signal intensity for both tissues analyzed in Figure 14.

In addition to increased detection of non-sialylated *m/z* peaks, combined PNGase F/sialidase treatment provides valuable structural information to identify and accurately analyze the spatial distribution of multi-fucosylated structures (Figure 14I-K). These fucoses are either attached to a GlcNAc on the outer arm of the glycan or at the core GlcNAc that attaches the glycan to the asparagine residue. N-glycans with two fucoses have a near identical theoretical mass to a +1 isotope of a structurally associated N-glycan with a single sialic acid. An example of this is found with the N-glycan structures F2A2G2 (*m/z* = 1955.6972) and A2G2S1 (*m/z* = 1954.6768). By utilizing

sialidase, the sialylated glycan's +1 isotope is eliminated, which allows for more accurate analysis of the multi-fucosylated N-glycan's spatial distribution. Particularly in cases where the expression of the multi-fucosylated N-glycan is low, sialidase utilization is critical to accurately analyzing the spatial distribution of these fucosylated N-glycans.

In addition to combined PNGase F/sialidase enzymatic treatment, other endoglycosidases can be included instead of PNGase F to provide additional structural information. Endo F3 preferentially targets core fucosylated N-glycans, with a different cleavage site as PNGase F. Therefore, we implemented combined Endo F3/sialidase treatment to target non-sialylated, core fucosylated structures specifically. This combined enzymatic digestion is expected to yield increased sensitivity compared to Endo F3 treatment alone. This methodology was implemented on two serial sections of an HCC tissue (Figure 15). The AUP of abundant non-sialylated core fucosylated structures was substantially increased with simultaneous sialidase treatment (Figure 15A-F). Meanwhile, Endo F3/sialidase treatment nearly completely eliminated m/z peaks for sialylated core fucosylated N-glycans (Figure 15G-H). These results mirror what is seen through PNGase F/sialidase treatment, with significantly higher signal intensity for non-sialylated glycans when treated with sialidase. Sialic acid removal improves detection of non-sialylated N-glycans with utilization of both PNGase F and Endo F3 as endoglycosidases to cleave glycan analytes from associated glycoproteins.

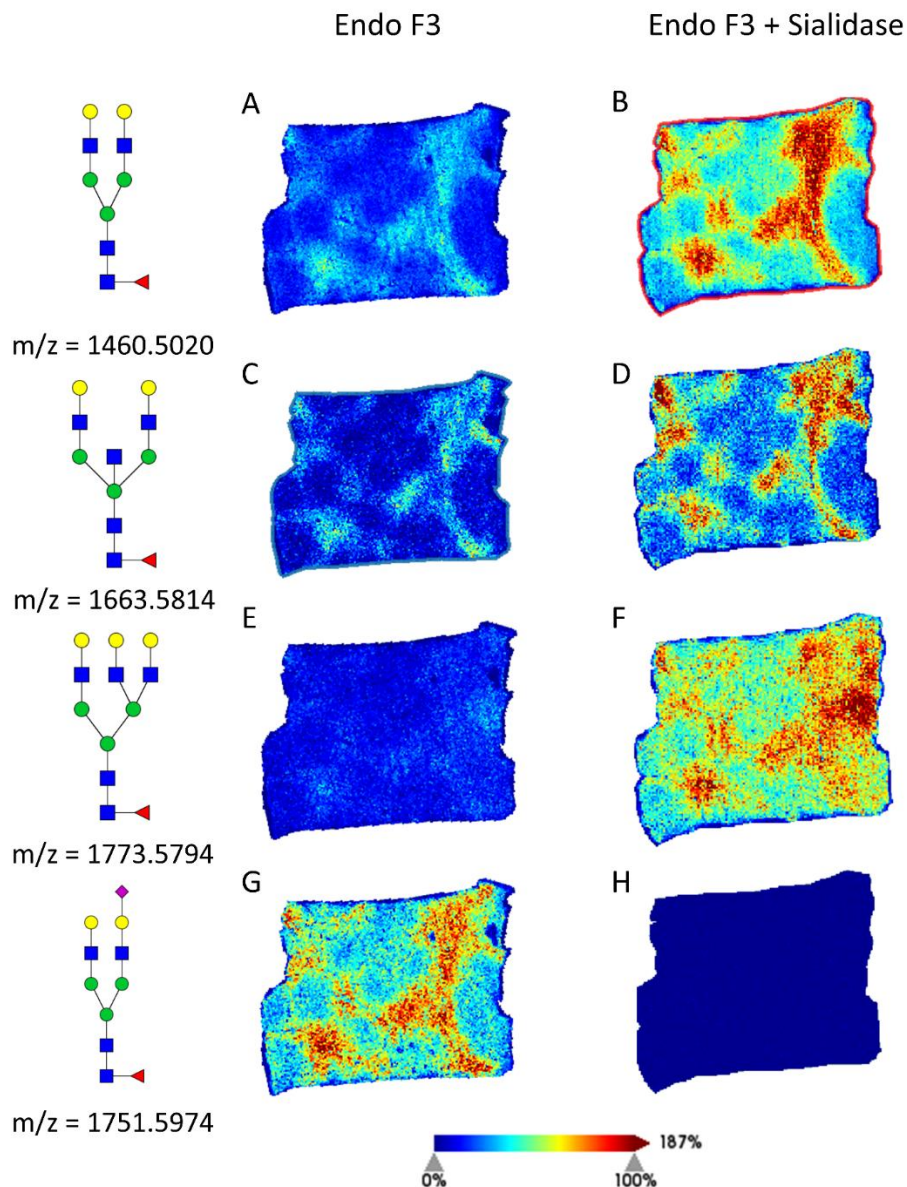


Figure 15. MALDI-IMS with Simultaneous Endo F3 and Sialidase Incubation Endo F3 and combined Endo F3/Sialidase Prime treatment is shown. Endo F3 cleaved m/z values represent the PNGase F cleaved structure with a different cleavage site, which results in a 349.1373 Da mass shift. (A-F): m/z images depicting increased signal intensity of abundant non-sialylated N-glycans with sialidase incorporation to Endo F3 N-glycan cleavage. (G-H): m/z images depicting the loss of signal intensity for a sialylated N-glycan with associated sialidase incubation.

4.3.3 UPLC Validation of On-Tissue Sialidase Activity

Based on MALDI-IMS data, combining PNGase F or Endo F3 with sialidase on-tissue led to substantially increased detection of non-sialylated N-glycans. This increase was larger than simply combining sialylated and non-sialylated m/z peaks, and thereby we hypothesized that endoglycosidase activity on-tissue was increased by removing sialic acids on the tissue. In order to demonstrate this, we prepared FFPE tissues under the same general workflow, but instead of applying MALDI matrix we collected enzymatically released N-glycans, labelled N-glycans with 2-AB, and analyzed via UPLC (Figure 16). N-glycans were collected from two serial sections of an HCC tissue, one sprayed with PNGase F/sialidase and one with only PNGase F. Results showed a 207.7% increase for A2G2, a 346.2% increase for FA2G2, and a 195.2% increase for A3G3. While some of the increase in peak intensity is attributable to desialylated glycans, summing the sialylated glycans detected on the tissue cannot account for such a large increase. This is supported both by the UPLC data and by a third serial section that was sprayed with only PNGase F and analyzed via MALDI-IMS (Figure 17). Figure 17 demonstrates a serial section of the tissue analyzed via UPLC, which was sprayed with only PNGase F. The sialylated glycoforms of the A2G2 base structure make up 145% of the AUP of A2G2 ($m/z = 1663.5814$), yet the peak for A2G2 increased by 207% in the UPLC experiment, suggesting that something beyond the combination of m/z peaks (through desialylation) is contributing to increased signal intensity. Overall, this result validates the increase to sensitivity that was observed through MALDI-IMS, which suggests that PNGase F/Endo F3 enzymatic efficiency is improved when incubated simultaneously

with sialidase. It is important to note that this improvement to enzyme efficiency occurs when the enzymes are applied directly onto a tissue through a sprayer, and it is not being discussed for in-solution enzyme digestions.

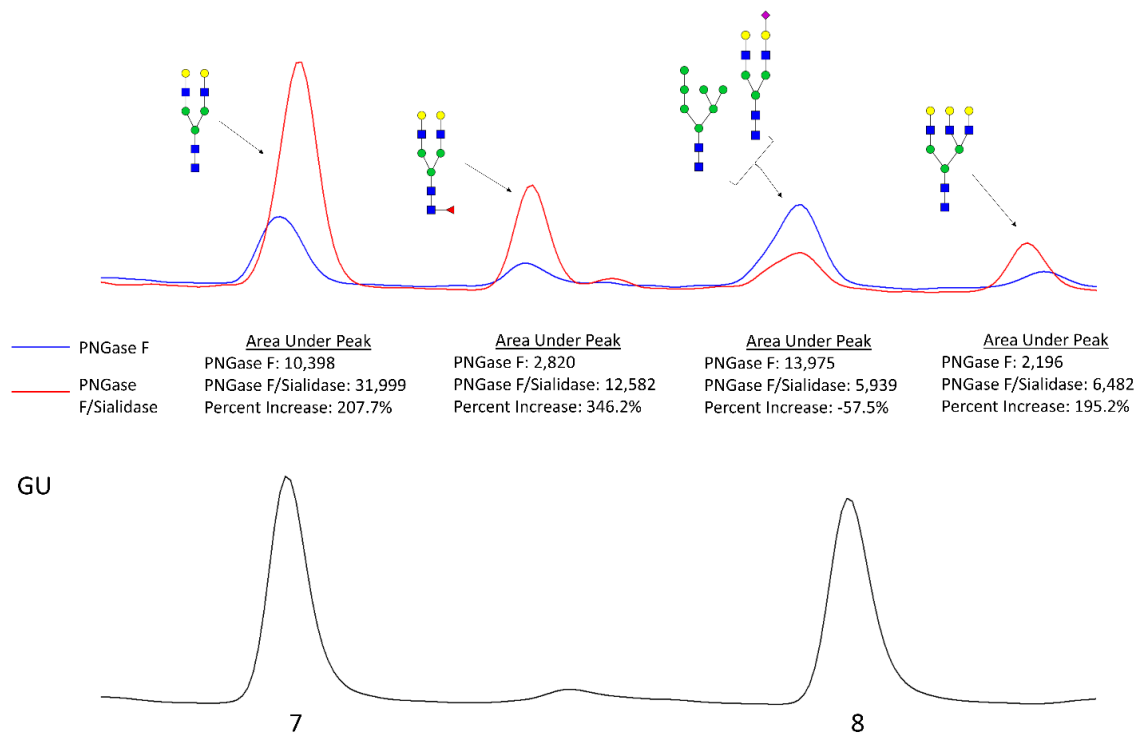


Figure 16. UPLC Analysis of On-Tissue Sialidase/PNGase F Activity. UPLC analysis of an on-tissue enzyme digestion of serial sections of an HCC tissue, one sprayed with PNGase F and one with PNGase F/Sialidase. There is a large increase to relative peak intensity for non-sialylated N-glycans when incubated simultaneously with sialidase. Peaks are labeled based on glucose units (GU), which are calculated based on retention time of a glucose homopolymer ladder. The peak at GU: 7.98 is labelled as a converging peak of A2G2S1 and M7.

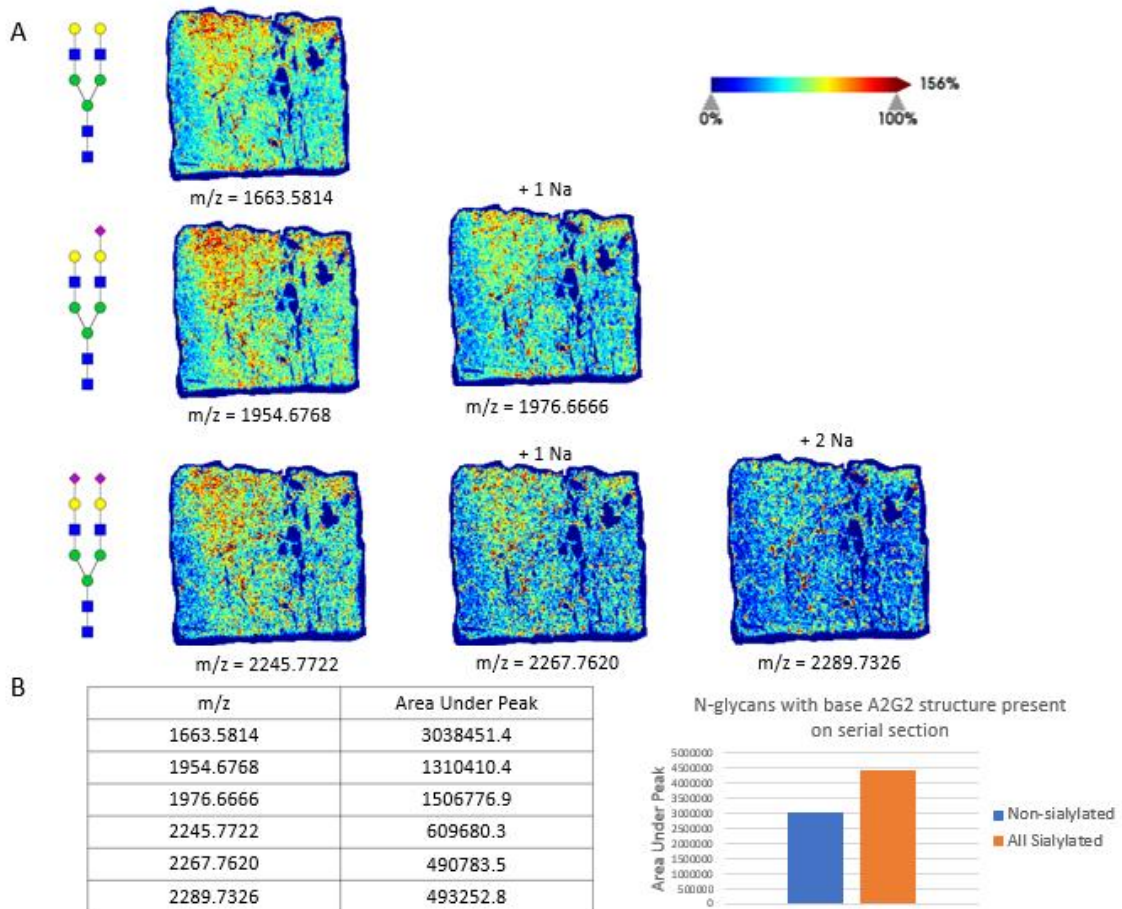


Figure 17. Analysis of Sialylated Glycoforms via MALDI-IMS. Through MALDI-IMS, we see a total AUP of 3038451.4 for the non-sialylated glycan A2G2 and a total AUP of 4410903.9 for all sialylated glycoforms of A2G2, which means that the combined sialylated glycoforms make up 145% of the non-sialylated glycoform's abundance. Through UPLC on serial sections sprayed with PNGase F and PNGase F/sialidase, A2G2 signal intensity increases by 207%, which suggests that increased enzymatic efficiency accounts for the difference between 145% and 207%.

4.4 Discussion

Sialylation of N-glycans is a common structural element that has been well-established to have significant biological function and to be relevant to the progression of some cancers.^{98,150,299,300} There are multiple methods of studying sialic acid expression, including through sialic acid binding lectins (siglecs), through UPLC/HPLC, and through MALDI-IMS.^{234,301} However, there are also biological cases in which N-glycan structural features other than sialic acid expression are correlated to disease progression, such as fucosylation and increased branching.¹⁴¹ In these cases, the expression of both sialylated and non-sialylated N-glycans adds complexity to analyses and decreases sensitivity for low abundant structures through the diversification of expressed structures, which is particularly relevant for MALDI-IMS techniques. Therefore, in this work we proposed the utilization of Sialidase Prime within an established MALDI-IMS N-glycan technique in order to enzymatically remove sialic acids from tissue N-glycans.

The most notable benefit of this work is found in increasing the signal intensity for non-sialylated glycan structures, which we found to be increased by more than expected due to offsetting detected signal of sialylated structures on serial sections. Both MALDI-IMS and UPLC analyses showed that signal intensity for non-sialylated N-glycans was increased by more than expected compared to the sum of removing all sialylated N-glycans. There were even increases to signal intensity of high mannose N-glycans, which would not be expected to be sialylated. This suggests that, when applied via a sprayer, PNGase F/Endo F3 enzyme efficiency is increased when simultaneously

incubated with sialidase. We expect this to be due to the steric benefit of removing sialic acids from the tissue surface, which allows for the endoglycosidase to have more access to the cleavage site. This hypothesis is supported by previous studies which have shown increased efficiency after removing large steric groups through previous enzyme activity directly on-tissue.²⁸⁹

In addition to increased signal intensity/endoglycosidase activity, sialidase utilization provides improved analysis of non-sialylated glycan structures that would otherwise overlap with sialylated N-glycan m/z peaks. Through the established MALDI-IMS N-glycan methodology, glycans with one sialic acid have a +1 isotope peak that often overlaps with the m/z peak of a glycan with the same base structure and two fucoses instead of the sialic acid. While FT-ICR mass spectrometers can in some cases have adequate resolving power to distinguish between these m/z peaks, the necessary resolving power depends on the relative expression of the two m/z peaks, and this would still be an unavoidable problem when using MALDI-TOF instruments.²⁷⁸ The interference of the +1 isotope can greatly affect the spatial distribution of the m/z image for the multi-fucosylated glycan, and makes analysis difficult when both glycans are expressed on a given tissue. For both this reason and the improved sensitivity, removing sialic acids with sialidase is a valuable experimental strategy for studies focusing on fucosylation.

While this technique can effectively improve non-sialylated N-glycan signal intensity and overall endoglycosidase activity, there are still elements of the protocol that can be further optimized. The choice of what sialidase to use can affect which sialic

acid linkages are preferentially cleaved. The enzyme Sialidase Prime used in this study, from the bacterium *Arthrobacter ureafaciens*, had comparable activity to commercially available α 2-3,6,8 neuraminidase and there was very little residual sialylated N-glycan signal detected via MALDI-IMS and UPLC. However, different sialidase enzymes could be used to alter which sialic acids are preferentially cleaved and more specific targeting of specific sialylated N-glycans could be done that way. Another optimization step would be to further multiplex enzymes for MALDI-IMS on the same tissue section, an example of which would be Endo F3/sialidase followed by PNGase F. This sialidase workflow could also be applied to what is already published regarding multiplexed N-glycan targeting with ECM targeting.²⁸⁹ However, sialidase utilization as shown here is an effective way to improve sensitivity of MALDI-IMS N-glycan analyses to specifically analyze N-glycan features unrelated to sialic acid expression.

Chapter 5: N-Glycomic Analysis of Patient-Matching Serum and Tissue

5.1 Introduction

N-glycosylation of glycoproteins is well known to be dysregulated in many cancer types, and it has shown to be aberrant on both tissue and within serum of HCC tumors.^{89,93,115,242,245,302} N-glycosylation directly on the tumor tissue surface has been demonstrated to directly impact both tumor proliferation and metastasis, including through the upregulation of growth factor receptor signaling.^{116,141,143,151} Previous work has shown an increase to branching and fucosylated N-glycans on HCC tissue when compared to cirrhotic or normal liver tissue.²⁴⁵ Additionally, aberrant N-glycosylation has been observed in HCC serum, which includes the differential N-glycosylation of several identified biomarkers of HCC such as AFP-L3.^{4,70} However, serum-based studies have largely relied on glycoproteins with known biomarker potential regardless of glycosylation, and studies to connect overall tissue N-glycosylation with serum N-glycosylation are lacking. Aberrant N-glycosylation is often heterogeneous between tumors, which has limited understanding in how tumor-specific N-glycosylation translates to the glycosylation of serum glycoproteins, even though many of those glycoproteins originate from the liver. In order to address this problem, recent advancements have been made for the analysis of N-glycans via MALDI-IMS for both tissue and serum samples, including for the specific analysis of serum glycoproteins through antibody arrays.^{219,235,288} If a better understanding could be developed in regard to how cancer-related N-glycosylation of both serum and tissue is related, better glycan-based serum biomarkers could be identified, which would significantly improve clinical outcomes overall.

N-glycosylation has been shown to be a key area for HCC biomarker development, although identification of a clinically implementable biomarker has been difficult. We hypothesize that this is largely due to the genetic and molecular heterogeneities of HCC, which can be accounted for through the incorporation of HCC subtypes. The Hoshida HCC classification system separates HCC into distinct subtypes based on genetic, clinical, and histological features, and previously published work has demonstrated differential N-glycosylation of subtyped HCC tissue.^{41,115} By accounting for some of the heterogeneity of HCC N-glycosylation through HCC subtyping, it becomes possible to identify more sensitive biomarkers which could work in a complementary fashion to diagnose each tumor subtype. However, in order to have clinical value it needs to be demonstrated that tumor subtype-specific N-glycosylation is detectable on serum glycoproteins, which is the next key step of this work in HCC biomarker development.

In order to address the disconnect between HCC-correlated tissue N-glycosylation and HCC-correlated serum N-glycosylation, we utilized a sample cohort of 23 patient-matching HCC tissues consisting of tumor tissue, background liver tissue, and serum samples. Additionally, there were 35 more patient-matching tumor and background liver tissue samples, which were used to validate previously published findings regarding tumor tissue-specific N-glycosylation. These samples are to be analyzed through previously published MALDI-IMS workflows, utilizing PNGase F Prime and Sialidase Prime to deglycosylate and desialylate N-glycans for mass spectrometry analysis.³⁰³ Glycomic analysis of patient-matching HCC tissue and serum samples is

novel, and provides important insight into the similarity between serum and tumor N-glycosylation. Establishing subtype-specific differences in N-glycosylation is critical to furthering the idea that subtyped-based analyses are valuable to the identification of serum-based cancer biomarkers.

5.2 Materials and Methods

5.2.1 Patient Tissues and Serum

FFPE tissue sections (5 μm thick) were made from 60 HCC tissue blocks and from 54 background liver tissue blocks that were obtained through surgical resection of HCC in the background of liver cirrhosis. Samples were provided from the UT Southwestern Medical Center under IRB 102010-051 to Dr. Amit Singal. Additionally, 23 serum samples were obtained, that matched to patients of obtained tissue samples. Serum samples were also provided from the UT Southwestern Medical Center under IRB 102010-051 to Dr. Amit Singal. All N-glycan imaging work at MUSC was performed under IRB Pro00079936. HCC tissues were subtyped according to the Hoshida classification system as done previously.^{42,256} All tissues were H&E stained following MALDI-IMS analysis and tumor regions were annotated by a liver pathologist. Patient characteristics of the 58 tumors able to be subtyped can be found in Table 10, and patient characteristics of the 23 cases with matching serum can be found in Table 11.

5.2.2 FFPE Tissue Preparation

HPLC grade methanol, ethanol, acetonitrile, xylene, and water were obtained from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid and α -cyano-4-hydroxycinnamic acid were obtained from Sigma-Aldrich (St. Louis, MO). Peptide-N-glycosidase F (PNGase F) Prime and Sialidase Prime was cloned, expressed, and purified in-house as previously described.^{219,303}

FFPE tissues were prepared according to a previously published protocol.^{219,234,270} Tissue Tack microscope slides were purchased from Polysciences Inc (Warrington, PA, Catalog No. 24216). Slides were taken through dewaxing and wash steps, before being processed by antigen retrieval in a Decloaking Chamber in 10 mM citraconic anhydride buffer, pH 3. PNGase F Prime and Sialidase Prime were simultaneously applied using a M5 TM-Sprayer Tissue MALDI Sample Preparation System (HTX Technologies, LLC), and enzymes were incubated for 2 hours at 37°C. MALDI matrix α -cyano-4-hydroxycinnamic acid (0.042 g CHCA in 6 mL 50% acetonitrile/49.9% water/0.1% TFA) was sprayed by the M5 TM-Sprayer.

5.2.3 N-Glycomic Total Serum Preparation

Serum samples were prepared for total serum N-glycan MALDI-IMS analysis through a previously published protocol.²⁸⁷ Hydrogel-coated slides (Nexterion Slide H) were obtained from Applied Microarrays (Tempe, AZ). Serum samples were spotted and immobilized on the slide surface, and then washed to remove salts and lipids from the sample. Each sample was spotted in triplicate. N-glycans were then enzymatically released through the same methodology of spraying/incubating PNGase F Prime/Sialidase Prime as tissue samples.

5.2.4 Serum Glycoprotein Antibody Array Preparation

Serum samples were prepared for glycoprotein-specific N-glycan MALDI-IMS analysis through a previously published antibody array protocol.^{288,304} Antibodies were

spotted onto nitrocellulose-coated microscope slides (Grace-Bio Labs, Bend, OR). Well chambers were mounted to the slide to create separated regions for serum incubation. Each region had 8 unique antibodies spotted at specified positions, with 16 total unique antibodies for analysis. Serum samples were incubated in duplicate in each of two antibody arrays. N-glycans were enzymatically removed via the same protocol as tissues and total serum analysis.

5.2.5 MALDI-IMS N-Glycan Imaging

Slides were imaged on a MALDI-TOF (timsTOF, Bruker Daltonics) mass spectrometer in positive ion mode (m/z 600-5000). For both tissue and serum, images were collected at a 150 μ M raster with 200 laser shots per pixel. Data was visualized and analyzed using SCiLS Lab 2022 (Bruker). Peaks were assigned to N-glycan structures utilizing a previously developed database with consideration for biosynthetic pathways of N-glycans. Not all N-glycans observed in tissue were also seen in serum samples. Proposed N-glycan structures and corresponding m/z values can be found in Table 9.

Theoretical <i>m/z</i>	Proposed N-Glycan Structure
933.3173	Hex3HexNAc2
1079.37	Hex3dHex1HexNAc2
1095.3729	Hex4HexNAc2
1136.3994	Hex3HexNAc3
1257.4235	Hex5HexNAc2
1282.4532	Hex3dHex1HexNAc3
1298.4481	Hex4HexNAc3
1339.4777	Hex3HexNAc4
1419.4788	Hex6HexNAc2
1444.507	Hex4dHex1HexNAc3
1460.5013	Hex5HexNAc3
1485.5347	Hex3dHex1HexNAc4
1501.5302	Hex4HexNAc4
1524.4973	Hex4dHex1HexNAc3 + SO4
1540.4579	Hex5HexNAc3 + SO4
1542.5544	Hex3HexNAc5
1546.4697	Hex4dHex1HexNAc3 + SO4 + 2Na
1562.5354	Hex5HexNAc3 + SO4 + 2Na
1581.5332	Hex7HexNAc2
1590.5657	Hex4dHex2HexNAc3
1606.5599	Hex5dHex1HexNAc3
1622.5548	Hex6HexNAc3
1647.5865	Hex4dHex1HexNAc4
1663.5814	Hex5HexNAc4
1688.6168	Hex3dHex1HexNAc5
1704.6112	Hex4HexNAc5
1724.5044	Hex6HexNAc3 + SO4 + 2Na
1727.5783	Hex4dHex1HexNAc4 + SO4
1743.5874	Hex8HexNAc2
1765.5264	Hex5HexNAc4 + SO4 + 2Na
1768.618	Hex6dHex1HexNAc3
1784.6096	Hex7HexNAc3
1809.6393	Hex5dHex1HexNAc4
1850.6699	Hex4dHex1HexNAc5
1866.6608	Hex5HexNAc5
1889.6274	Hex5dHex1HexNAc4 + SO4
1905.6339	Hex9HexNAc2
1911.5863	Hex5dHex1HexNAc4 + SO4 + 2Na
1946.5468	Hex8HexNAc3
1955.6994	Hex5dHex2HexNAc4
1996.7205	Hex4dHex2HexNAc5
2010.7038	Hex7dHex1HexNAc3 + SO4
2012.7187	Hex5dHex1HexNAc5

2028.7132	Hex6HexNAc5
2035.6896	Hex5dHex2HexNAc4 +SO4
2067.6888	Hex10HexNAc2
2092.5944	Hex5dHex1HexNAc5 + SO4
2092.7048	Hex8dHex1HexNAc3
2101.7536	Hex5dHex3HexNAc4
2158.7721	Hex5dHex2HexNAc5
2174.7737	Hex6dHex1HexNAc5
2215.798	Hex5dHex1HexNAc6
2231.793	Hex6HexNAc6
2254.7511	Hex6dHex1HexNAc5 + SO4
2276.7146	Hex6dHex1HexNAc5 + 1SO4 + 2Na
2304.8322	Hex5dHex3HexNAc5
2320.8294	Hex6dHex2HexNAc5
2361.853	Hex5dHex2HexNAc6
2377.8509	Hex6dHex1HexNAc6
2393.8475	Hex7HexNAc6
2434.8734	Hex6HexNAc7
2466.8881	Hex6dHex3HexNAc5
2523.9092	Hex6dHex2HexNAc6
2539.907	Hex7dHex1HexNAc6
2580.9316	Hex6dHex1HexNAc7
2596.9272	Hex7HexNAc7
2669.9708	Hex6dHex3HexNAc6
2685.9638	Hex7dHex2HexNAc6
2742.9847	Hex7dHex1HexNAc7
2758.9826	Hex8HexNAc7
2800.0079	Hex5dHex5HexNAc6
2816.0246	Hex6dHex4HexNAc6
2832.0254	Hex7dHex3HexNAc6
2905.0399	Hex8dHex1HexNAc7
2978.0824	Hex7dHex4HexNAc6
3051.0938	Hex8dHex2HexNAc7
3124.1178	Hex9HexNAc8
3197.1611	Hex8dHex3HexNAc7
3270.1737	Hex9dHex1HexNAc8
3416.2318	Hex9dHex2HexNAc8
3635.3078	Hex10dHex1HexNAc9

Table 9. N-Glycan Peak List. Each *m/z* peak that was detected and associated with a N-glycan is listed. Not all N-glycans were detected in a given tissue/serum sample. Analyses comparing tissue and serum utilized a consolidated peak list of only N-glycans detected in both sample types.

5.2.6 Statistical Data Analysis

After MALDI-IMS experiments, statistical analyses were done to analyze both tissue and serum, including for the differentiation of HCC subtypes. For tissue: tissue or serum: serum analyses, all N-glycans of the peak list were percentiled to yield a relative quantification, and these relative quantifications were normalized for tissue: serum analyses. General linear models were developed and optimized utilizing linear regression, which was used to identify individual glycans that could significantly differentiate tumor from non-tumor along with groupings of multiple N-glycans to form a classification model. Feature selection for classification models was based on classification ability of individual N-glycans. These classification models were internally validated using leave one out cross validation. Individual glycans that were significantly different between subtypes were identified using pairwise t-tests, p-value less than 0.05 considered significant. Pearson correlations were performed to identify N-glycans that were significantly correlated between tumor and serum, p-values less than 0.05 considered significant.

5.3 Results

5.3.1 Demographic and Clinical Information of Sample Cohort

The set of samples being analyzed in this work is both unique and incredibly valuable for a variety of reasons. For N-glycomic analysis, there are 66 tumor FFPE tissue blocks, originating from 60 patient cases. Of these 60 cases, 54 of them include a patient-matching background liver FFPE tissue block, which is a tissue section that is non-tumorous but originating from a tumor-containing liver. These 60 cases were also subtyped according to the Hoshida classification system, with 2 patient samples unable to be subtyped due to tumor size. Overall, this grouping of tissue samples allows for a powerful analysis of the N-glycomic transformation that takes place from non-tumorous liver tissue to HCC tissue within the same patient, which allows for control over demographic/ethnicity-based differences in N-glycosylation. The main drawback is in application to a biomarker discovery study, where it would be better to be able to distinguish between liver tissue with no tumor present and HCC tissue, although evidence that N-glycosylation is altered in tumor-adjacent tissue compared to normal/cirrhotic tissue is limited. A summary of demographic and clinical information on these tissue samples is included in Table 10.

	S1 Tumors (n = 16)	S2 Tumors (n = 13)	S3 Tumors (n = 29)
Sex	11 Male/ 5 Female	11 Male/ 2 Female	20 Male/ 9 Female
Age	56.6 (± 10.8)	61.2 (± 7.0)	59.2 (± 10.6)
Race/Ethnicity	11 Black or African American/ 4 White Non-Hispanic/ 1 White Hispanic	9 Black or African American/ 3 White Non-Hispanic/ 2 Asian	13 Black or African American/ 9 White Non-Hispanic/ 7 White-Hispanic
Etiology	13 HCV/ 2 Cardiogenic/ 1 Alcohol-induced	12 HCV/ 1 Alcohol-induced	19 HCV/ 4 Alcohol-induced/ 1 NASH/ 1 HBV/ 4 Unknown
Child-Pugh	13 A/ 3 B	12 A/ 1 B	27 A/ 2 B
Number of Lesions	16 with 1 Lesion	12 with 1 Lesion/ 1 with 3 Lesions	28 with 1 Lesion/ 1 with 2 Lesions
Tumor Burden	2.6 cm (± 0.9)	3.3 cm (± 1.0)	4.1 cm (± 3.5)
BCLC Stage	1 Stage 0/ 15 Stage A	12 Stage A/ 1 Stage B	2 Stage 0/ 26 Stage A/ 1 Stage B
AFP Level (ng/mL)	9 > 20 ng/mL (56.3%)	7 > 20 ng/mL (53.8%)	8 > 20 ng/mL (27.5%)
Recurrence/ Outcome	5 with Recurrence/ 2 Unknown/ 9 No Recurrence	3 with Recurrence/ 1 Unknown/ 9 No Recurrence	15 with Recurrence/ 2 Unknown/ 12 with No Recurrence

Table 10. Demographic/Clinical Information of Tissue Sample Cohort.

Sex/age of the sample cohort is well representative of the incidence rates of HCC, which includes approximately a 2:1 higher incidence rate in men. This cohort is primarily made up of Black/African American patients, although samples sizes are too small to do statistically significant analyses based on race. Similarly, this cohort primarily is made up of tumors with an HCV etiology, although samples sizes are too small to do statistically significant analyses based on etiology. These samples were primarily early-stage tumors with only one small lesion, which is due to the method of sample acquisition for these studies being resection. Interestingly, both S1 and S2 tumors appear to have similar rates of AFP positive tumors within this cohort, although the Hoshida classification system lists S2 tumors as AFP-positive. This reinforces the idea that these subtypes are utilizing a significant amount of classification data, and there is still substantial variation within each grouping. Although recurrence data is provided, due to small sample sizes and an inability to distinguish early and late recurrence there will not be statistical analyses on recurrence as a part of this study.

In addition to the patient-matching tumor and background liver tissue samples, 23 serum samples were included in this cohort, all of which match to corresponding tissue samples. This is an incredibly unique set of samples, and allows for analysis into how tumor-specific N-glycosylation translates into serum N-glycosylation. Considering that these are patient-matching samples, subtype-based analyses can also be performed based on serum N-glycosylation. However, this becomes a problem due to a low number of S2 serum samples, leading to analyses attempting to distinguish S3 serum samples from a combination of S1/S2 serum samples. Combining spatial N-glycomics of

tumor tissue samples with serum N-glycomics of patient-matching samples is novel and an exciting opportunity in the field of liver cancer biomarker research. A breakdown of clinical/demographic characteristics of the cases that include serum samples is included in Table 11.

	Tissue-matching serum samples (n = 23)
Sex	14 Male/ 9 Female
Age	61.3 (\pm 5.0)
Race/Ethnicity	15 Black or African American/ 5 White Non-Hispanic/ 3 White Hispanic
Hoshida Subtype	6 S1/ 3 S2/ 14 S3
Etiology	17 HCV/ 5 Alcohol-induced/ 1 NASH
Child-Pugh	23 A
Number of Lesions	22 with 1 Lesion/ 1 with 2 Lesions
Tumor Burden	4.3 cm (\pm 3.7)
BCLC Stage	1 Stage 0/ 21 Stage A/ 1 Stage B
AFP Level (ng/mL)	7 > 20 ng/mL (56.3%)
Recurrence/ Outcome	12 with Recurrence/ 11 No Recurrence

Table 11. Demographic/Clinical Information of Serum Sample Cohort.

5.3.2 N-Glycosylation of HCC Compared to Background Liver

The first analysis done with this sample cohort was a direct comparison between tumor N-glycosylation and background liver N-glycosylation, with the aim to validate that there are detectable differences in N-glycosylation between those two tissue types. Each patient case in this analysis contains a tissue block containing HCC tumor, and a background liver tissue block containing no tumor, which was distant from the tumor but still removed during resection. This is a different analysis than done in Chapter 3, which relied on directly adjacent cirrhotic/normal tissue. When doing MALDI-IMS N-glycan imaging, tumor tissues contain distinct tumor and non-tumor regions, while background liver tissues are relatively uniform in N-glycosylation with fewer histological features of interest aside from fibrosis. This difference in images is displayed in Figure 18. Tumor regions are still annotated by a pathologist, but the remainder of the tissue on the tumor tissue sections is not incorporated into the analysis. Background liver tissue sections were also annotated by a pathologist, with regions of significant fibrosis excluded from analysis. The relative expression of each N-glycan of the m/z list is analyzed and compared between the tumor and the corresponding background liver tissue. This allows for a direct comparison between different MALDI-IMS runs of the tumor and background liver.

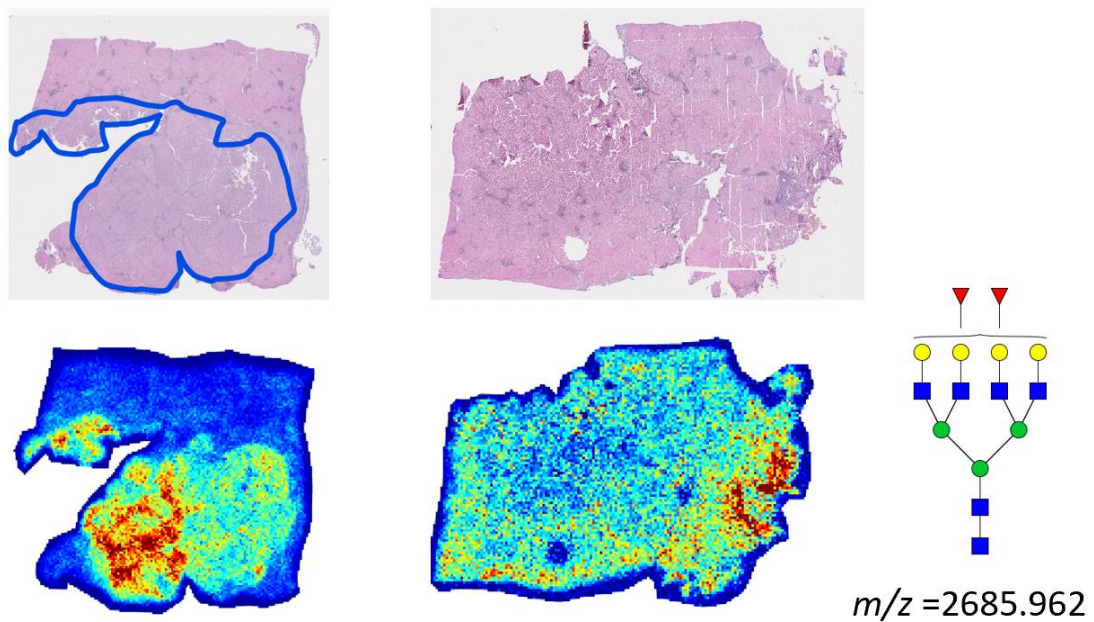


Figure 18. Example m/z Images of Tumor and Background Liver Tissue. Displayed is an example of a tumor tissue section (left) and representative m/z image along with a patient-matched background liver tissue section (right) and representative m/z image.

Previous findings have demonstrated that, without taking tumor subtypes into consideration, N-glycosylation of HCC can be distinguished from non-HCC, but lacks the sensitivity to identify a singular N-glycan structure as a biomarker candidate. Within this sample set, that premise holds true, as it is possible to discriminate tumor from background liver tissue, but no single N-glycan structure is able to effectively do so alone. However, multivariate statistical approaches allow for the development of classification models combining multiple N-glycans, and this yields good results for the on-tissue discrimination of tumor and background liver. An optimized model is shown in Figure 19, with an AUC of 0.9842 and internal validation through leave one out cross validation (LOOCV). Interestingly, this model includes an incomplete biantennary glycan, a hybrid N-glycan (increased in non-HCC), and a tetra-antennary non-fucosylated N-glycan. These are not the large, fucosylated N-glycans expected to be increased in HCC, but they still combine to form a very accurate HCC model in this sample set. Overall, this work validates what has previously been demonstrated both in this dissertation and in literature, which is that N-glycosylation on tissue is significantly different between HCC and non-HCC liver tissue.

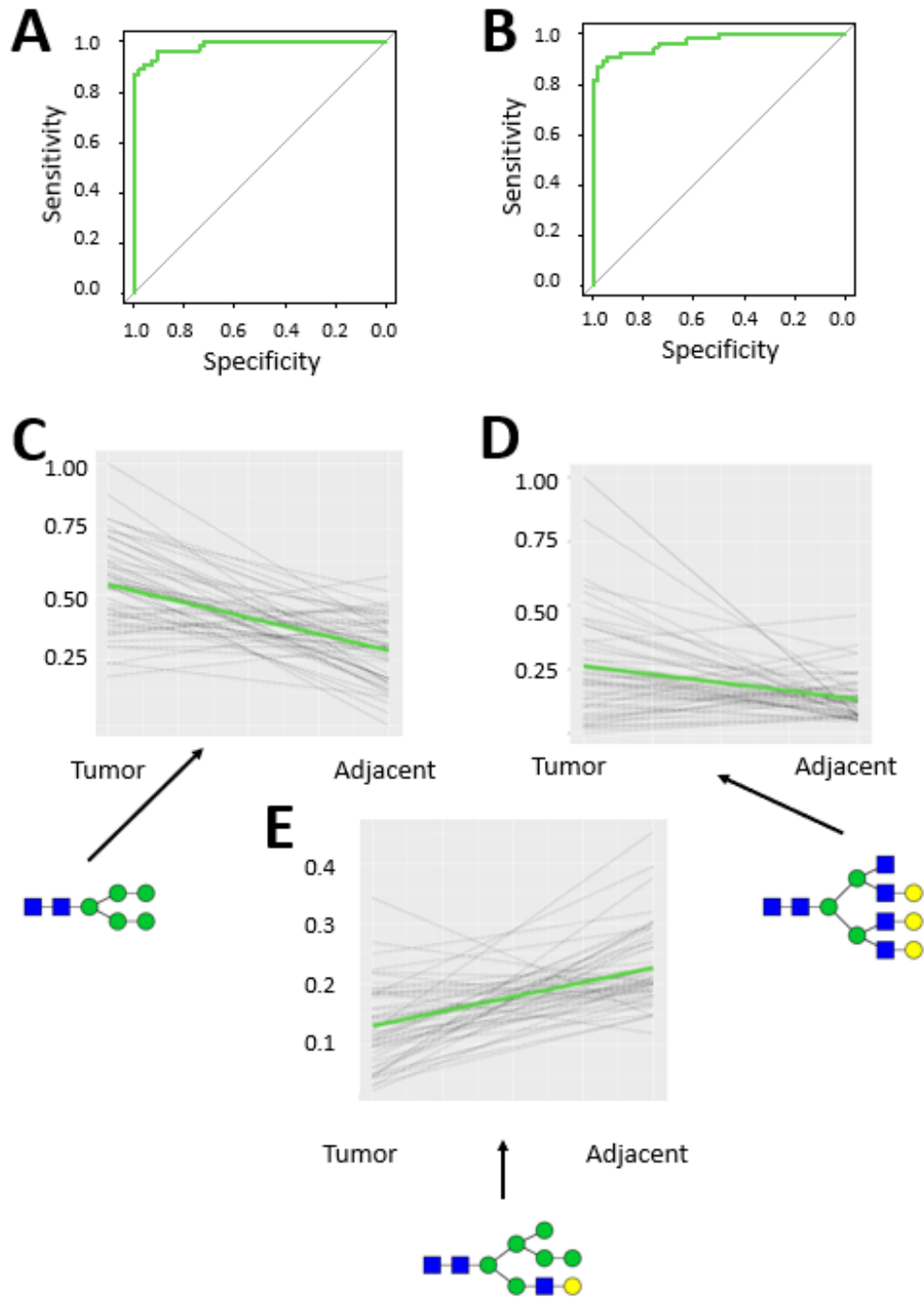


Figure 19. Classification Model for Tumor and Background Liver Tissue. Displayed is the optimized classification model to distinguish tumor from background liver tissue in this sample set. (A): The ROC curve for the model is shown, with an AUC = 0.9842, SE = 0.0084. (B): The ROC curve for the model following leave one out cross validation, with an AUC = 0.9698, SE = 0.0142. (C-E): On a normalized scale, the relative expression of each glycan of the model is shown, with patient-matched pairs between tumor and background liver tissue. (C): $m/z = 1257.424$, (D): $m/z = 1784.610$, (E): $m/z = 2231.793$.

5.3.3 Distinguishing HCC Subtypes in Tissue

With this larger sample cohort, it became possible to do analyses to validate what was previously seen in tissue regarding the differentiation of the N-glycosylation of subtyped HCC tumors. Within this sample set there are 16 S1 tumors, 13 S2 tumors, and 29 S3 tumors, which allows for robust statistical analysis to attempt to differentiate the tumor subtypes based solely on N-glycosylation. As expected, none of the analyzed N-glycans are capable of adequately discriminating the subtypes when acting as individual biomarkers, as glycan heterogeneity is still too substantial. Within a general linear model to predict all three subtypes, no individual glycan within the tumor region has an overall accuracy greater than 0.6. The individual N-glycans within the tumor region best suited to distinguish between all three of the subtypes are shown in Figure 20A. In addition, it is important to note that there are also no N-glycans in the matching background liver tissue regions capable of adequately predicting tumor subtypes. Within a general linear model to predict all three subtypes, no individual glycan within the background liver region has an overall accuracy greater than 0.55. The individual N-glycans within the background liver region best suited to distinguish between all three of the subtypes are shown in Figure 20B. Taken together, this suggests that there is more subtype-predictive capabilities in the tumor than in the background liver tissue, although neither tissue type contains a single N-glycan capable of discriminating all three subtypes adequately.

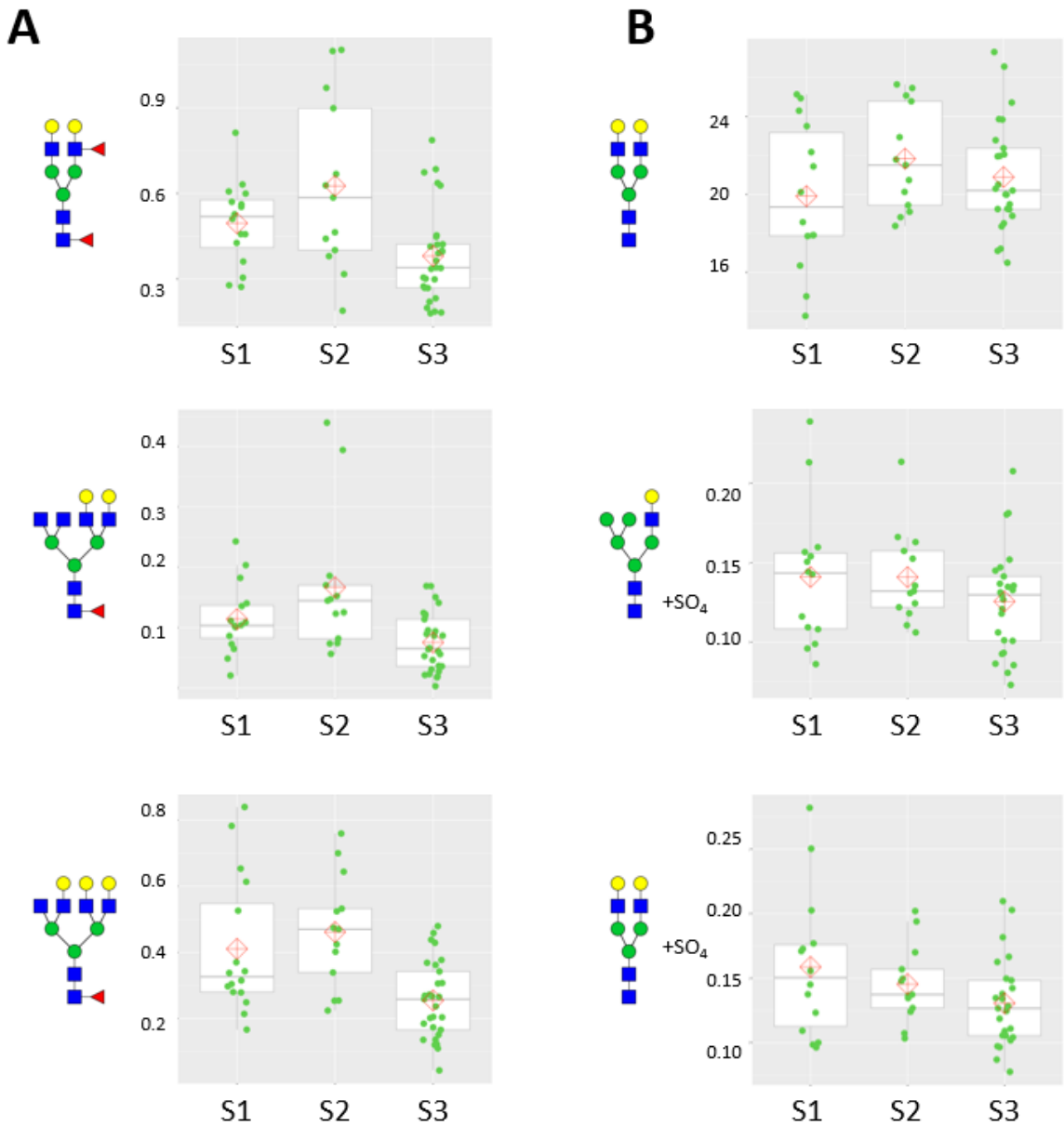


Figure 20. Best Individual N-glycans for Subtype Discrimination in Tumor and Background Liver. The best performing individual N-glycans, ranked by average AUC for discriminating each of the three subtypes, are displayed for tumor N-glycans (A) and background liver N-glycans (B). The y-axis is relative intensity of N-glycan expression. For tumor N-glycans: $m/z = 1955.699$; AUC = 0.709 (top), $m/z = 2215.798$; AUC = 0.698 (middle), $m/z = 2377.851$; AUC = 0.716 (bottom). For background liver N-glycans: $m/z = 1663.581$; AUC = 0.605 (top), $m/z = 1724.504$; AUC = 0.598 (middle), $m/z = 1765.526$; AUC = 0.612 (bottom).

While utilizing a single N-glycan to discriminate between all three subtypes has shown to be inadequate for biomarker development, there are interesting results when attempting to use single N-glycans to distinguish between two of the subtypes. Previous results demonstrated that when compared to directly adjacent non-HCC tissue, S1 and S2 were the most dysregulated, particularly in regard to fucosylation. Interestingly, there was much less distinction between S1 and S2 tumors in this cohort, and there was significant distinction between S1/S2 and S3 tumors. Utilizing a pairwise t-test to compare each subtype's expression of each N-glycan, there is only one N-glycan that displays significant statistical difference ($p < 0.05$) between S1 and S2 tumors, which is $m/z = 1727.578$; F1A2G1 + SO₄. Meanwhile, there are 11 N-glycans that are significantly different between S1 and S3 tumors, and there are 18 N-glycans that are significantly different between S2 and S3 tumors. The one N-glycan that statistically significantly differentiates S1 and S2 tumors is a low abundant sulfated, fucosylated N-glycan. Of the 11 N-glycans that differentiate S1 and S3 tumors, six of them are fucosylated, all of which are increased in S1 compared to S3 tumors. Of the 18 N-glycans that differentiate S2 and S3 tumors, 11 of them are fucosylated, all of which are increased in S2 compared to S3 tumors. The N-glycan that is significantly different between S1 and S2 is shown in Figure 21A, the three most significantly different N-glycans between S1 and S3 are shown in Figure 21B, and the three most significantly different N-glycans between S2 and S3 are shown in Figure 21C. Taken together, these results suggest that the N-glycosylation, and particularly fucosylation, of S3 tumors is much more distinct from S1 and S2 tumors than each of those subtypes are from each other.

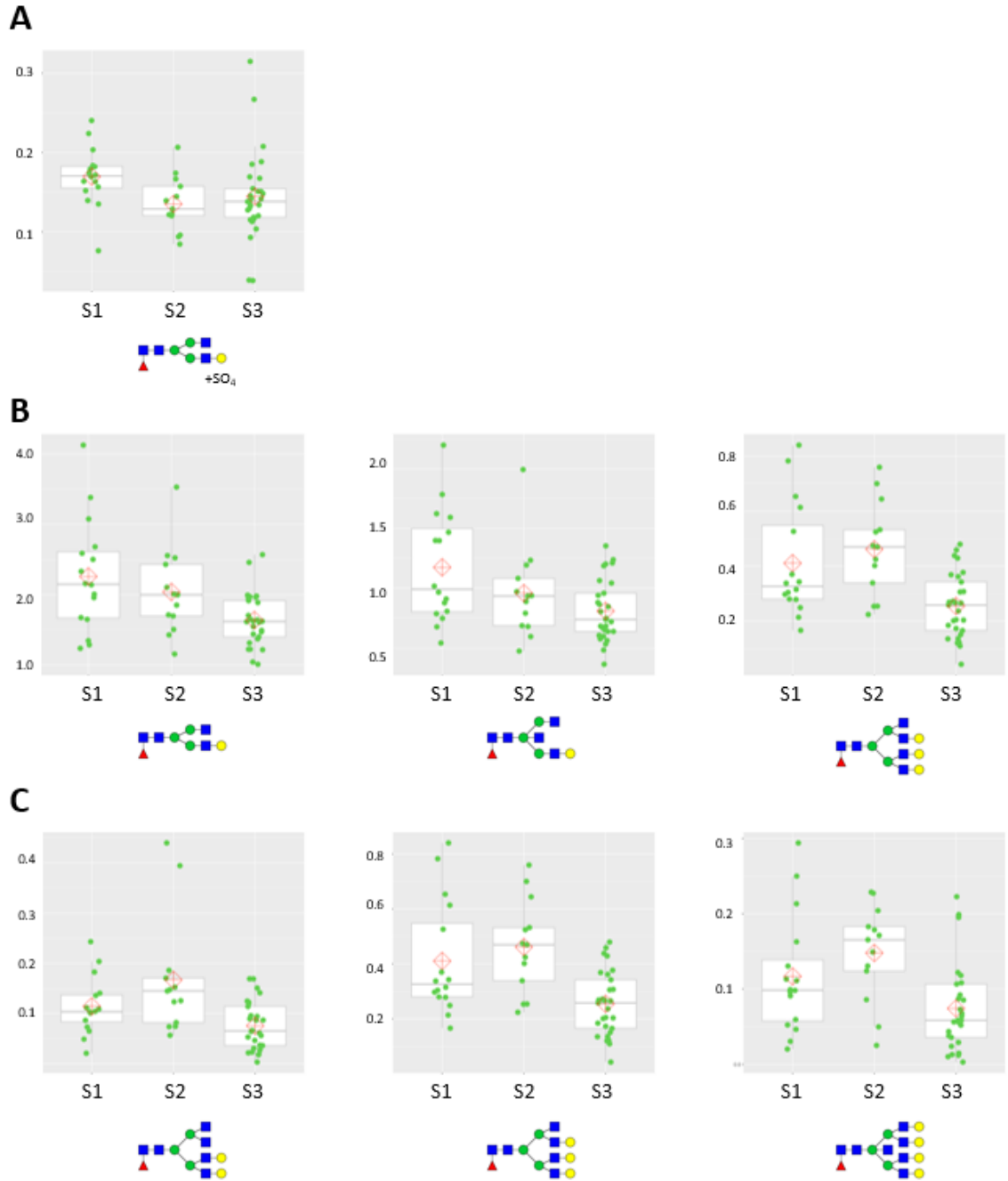


Figure 21. N-Glycans to Distinguish Each Subtype Pairwise. Utilizing pairwise statistical testing, the N-glycans that are significantly different between each subtype were determined. (A): The only N-glycan significantly different between S1 and S2. $m/z = 1727.578$. (B): The three most statistically significant N-glycans (of 11 total) to distinguish S1 and S3. $m/z = 1647.567$, $m/z = 1850.670$, $m/z = 2377.851$. (C): The three most statistically significant N-glycans (of 18 total) to distinguish S2 and S3. $m/z = 2215.798$, $m/z = 2377.851$, $m/z = 2742.985$.

While there is no individual N-glycan that could be adequately used as a standalone biomarker for the categorization of HCC subtypes, it is possible to construct an algorithm made up of multiple N-glycans that is highly sensitive for the classification of HCC subtypes. Figure 22 displays the most optimized linear regression model out of this N-glycomics data set for the classification of tumors into each of the three HCC subtypes. In Figure 22A, the ROC curve of this model is shown, demonstrating quality biomarker characteristics for the identification of all three HCC subtypes. The structure of each of the N-glycans utilized within the model are shown in Figure 22B. The relative importance of each N-glycan to the model is as follows: $m/z = 1606.560$: 36.069; $m/z = 1727.578$: 97.580; $m/z = 1768.618$: 104.075; $m/z = 2028.713$: 1.098; $m/z = 2158.772$: 74.672; $m/z = 2377.851$: 20.281; $m/z = 2580.932$: 140.373. Six out of the seven N-glycans are fucosylated, demonstrating once again the importance of fucosylation in the differentiation of HCC subtypes. Figure 22C displays a confusion matrix table for this classification model, demonstrating its overall accuracy to be 75.86%. The model correctly categorized 10/16 S1 tumors, 9/13 S2 tumors, and 25/29 S3 tumors, suggesting that it is particularly effective in correctly classifying S3 tumors. Lastly, Figure 22D displays the sensitivity, specificity, positive predictive value, and negative predictive value for each subtype. The AUC for S1 is 0.896, the AUC for S2 is 0.923, the AUC for S3 is 0.900, and the macro-AUC is 0.903. Overall, this is a highly effective model in utilizing N-glycosylation of tumor tissue samples within this sample set to categorize HCC subtypes, and external validation will be required to assess its potential clinical utility.

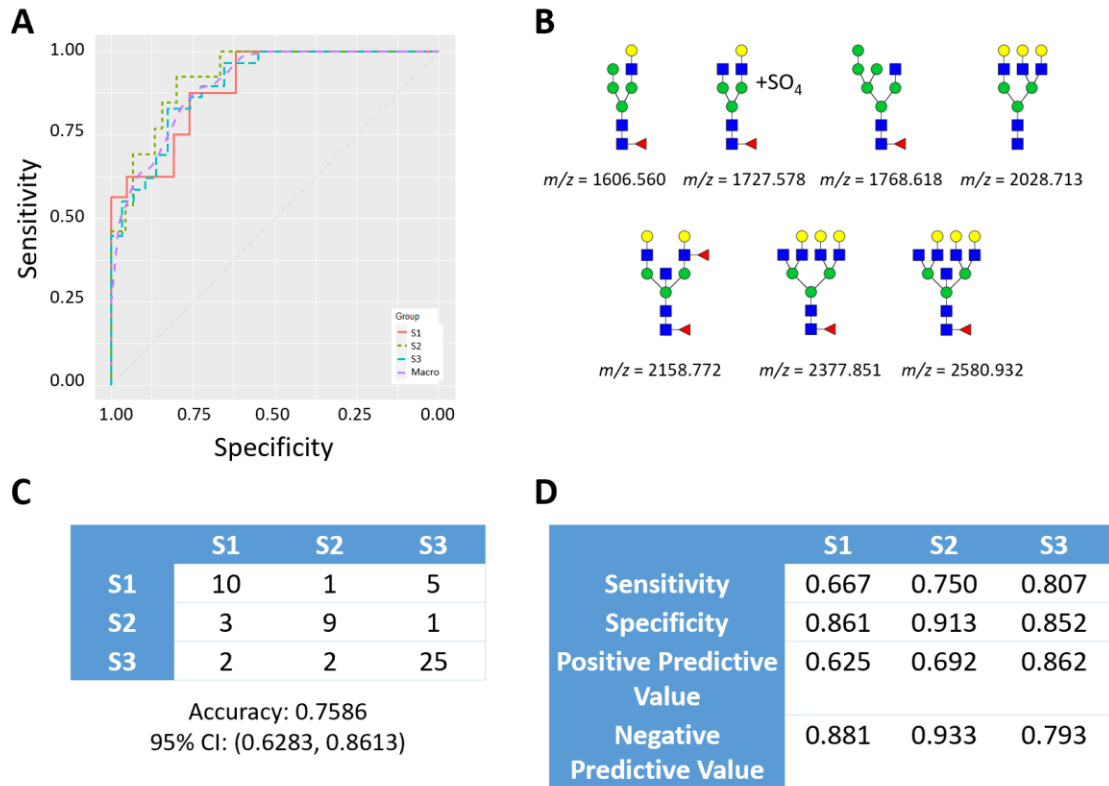


Figure 22. Optimized Model to Classify HCC Subtypes Utilizing N-Glycosylation.

Displayed are the characteristics of the linear regression model most optimized to classify HCC subtypes from this data set. (A): The ROC curve displaying sensitivity/specificity for all three subtypes. (B): The seven N-glycans included within this model. (C): A confusion matrix table demonstrating the accuracy of the model. (D): Sensitivity, specificity, positive predictive value, and negative predictive value of the model for the subtype.

5.3.4 Comparison of Tissue and Serum N-Glycosylation

After a thorough analysis of tissue compared to background liver tissue and subtype-based analyses only within tissue samples, the next step is to analyze the translation of N-glycan findings to serum. This sample set includes 23 patient-matching serum samples to tissue samples included in the previous analyses. The first aim was to analyze the correlation between N-glycosylation of the pathologist-annotated tumor region and total serum N-glycosylation through MALDI-IMS. Serum N-glycosylation of all glycoproteins present in serum was analyzed, suggesting that correlation will likely increase with more specific analyses on liver-originating serum glycoproteins. Pearson correlations were determined for all N-glycans in the m/z peak list, which determined the correlation between total serum N-glycosylation and tumor-specific tissue N-glycosylation. Results showed that 18/73 (24.7%) N-glycans are significantly correlated between tumor and total N-glycosylated serum. All of these were positively correlated, which suggests that HCC tumor and serum N-glycosylation are largely positively correlated, and thereby N-glycosylation trends on tumor tissue are likely to be represented in serum. Figure 23A shows the regression lines of all N-glycans that are significantly positively correlated, Figure 23B shows the remaining, non-significantly correlated N-glycans, and Figure 23C identifies the N-glycans that are significantly positively correlated between serum and tissue.

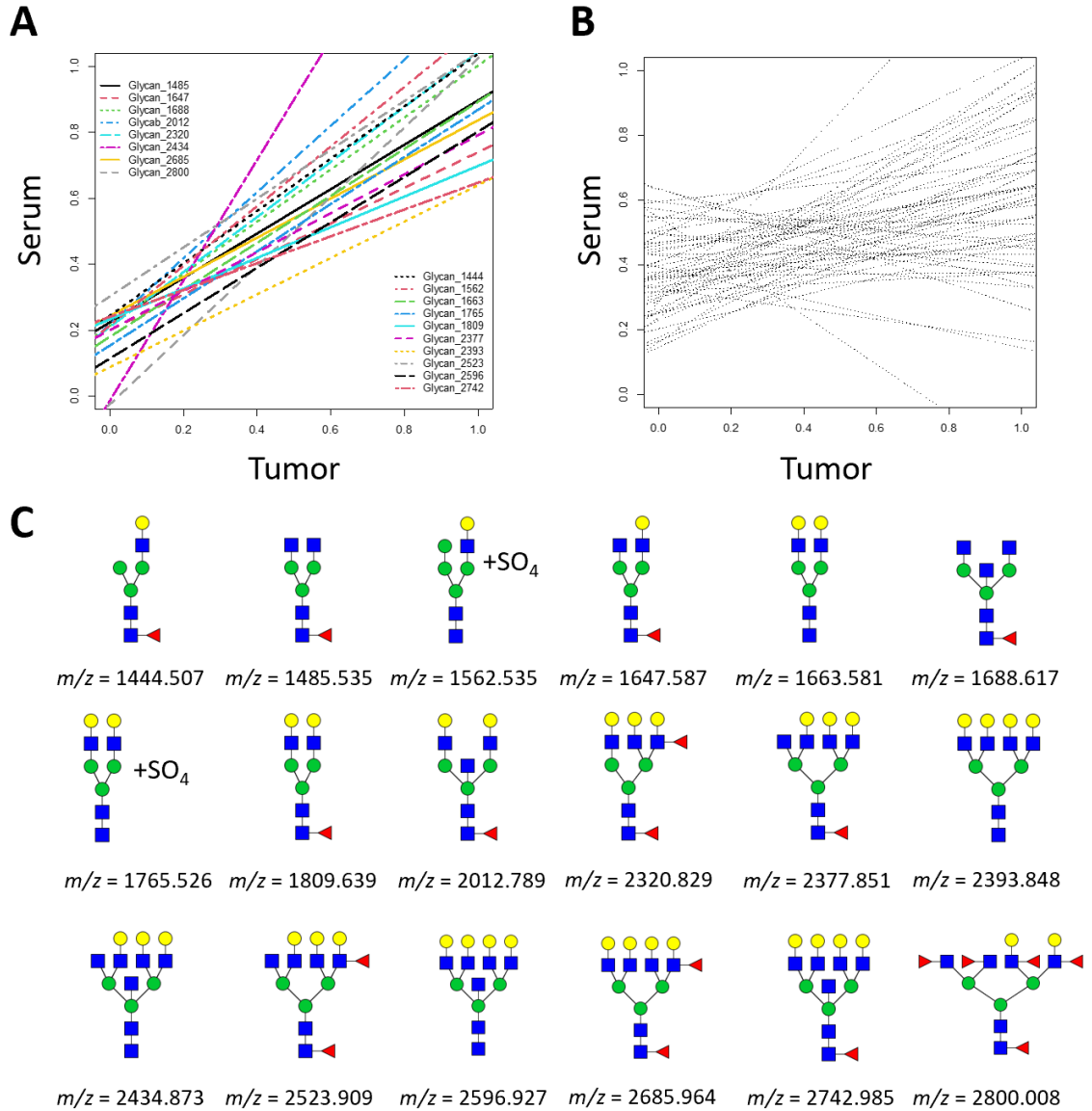


Figure 23. Pearson Correlations of Tumor and Serum N-Glycans. Displayed are the results from analyzing statistically significant correlations of the expression of each N-glycan in matching serum/tissue. (A): The regression lines of the 18 N-glycans that are statistically significantly correlated are shown. ($p < 0.05$) (B): The regression lines of the remaining N-glycans, none of which are statistically significant, are shown. (C): The 18 N-glycans significantly correlated between tumor and serum are shown.

5.3.5 Subtyped Serum Analysis

After determining that there is statistically significant correlation between tumor and serum N-glycosylation, it is important to analyze whether subtype-dependent tumor N-glycosylation is also evident in serum glycoproteins. Unfortunately, within the sample set of 23 patient-matching samples, there are only 3 S2 serum samples, which is insufficient to statistically analyze the distinction of all three subtypes from serum N-glycosylation. However, S1/S2 can be combined to compare to S3 for the serum samples, which is an important analysis as S3 tumors have been shown to have clearly distinct N-glycosylation from S1/S2 on tissue samples. To begin, a general linear model was optimized to distinguish S1/S2 from S3 using total serum N-glycosylation, which is agnostic to the originating glycoprotein. Results demonstrated that the utilization of two N-glycans, $m/z = 1419.479$; M6 and $m/z = 1647.587$; FA2G1, leads to a model with adequate sensitivity in differentiating S1/S2 from S3 serum samples. The details of this classification model are shown in Figure 24. This model achieved an AUC of 0.818 with a SE of 0.0982, and underwent LOOCV with an AUC of 0.714. While not sensitive enough to be clinically applicable, this is a promising model as it relies upon total serum N-glycosylation, and many of the included glycoproteins do not originate from the tumor or the liver. With a focus on more specific glycoproteins of interest, we'd expect to see improvement in subtype classification.

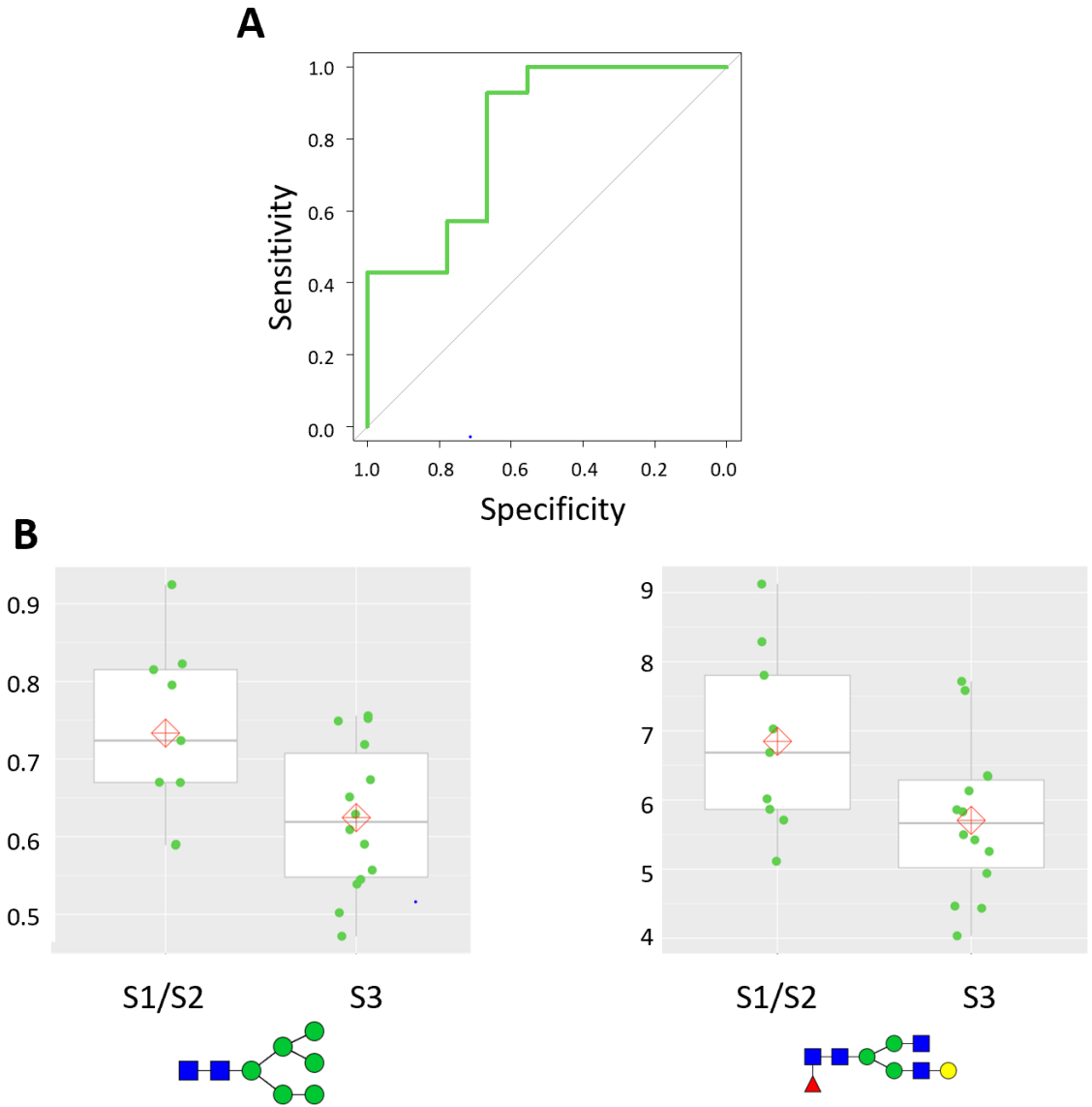
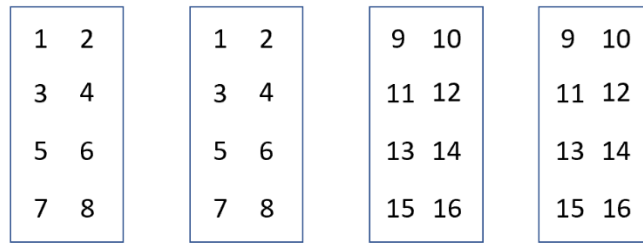


Figure 24. Total Serum N-Glycan Model to Differentiate Subtypes. (A): The ROC curve used to differentiate S1/S2 from S3 serum using total serum N-glycosylation. This model includes an AUC of 0.818. (B): The two N-glycans included within this model are $m/z = 1419.479$ (left), and $m/z = 1647.587$ (right).

In order to specifically analyze individual glycoproteins, we utilized a previously published MALDI-IMS antibody array workflow.²⁸⁸ This allows for the specific capture, and subsequent N-glycan analysis, of glycoproteins of interest, which in this case includes liver-originating glycoproteins and serum glycoproteins previously linked to cancer diagnostics. Figure 25 outlines all of the serum glycoprotein-specific antibodies selected for use in the antibody array. Included are a mix of abundant serum glycoproteins, glycoproteins known to be liver-secreted, and glycoproteins previously linked to HCC. Each antibody included in the panel was validated for specificity to its target glycoprotein. Thereby, this antibody array will allow for the analysis of HCC in subtyped serum samples in a more sensitive manner than total serum analyses.



- | | |
|--|--|
| <ol style="list-style-type: none"> 1. Alpha-1 antitrypsin (A1AT) 2. Alpha-1-B Glycoprotein (A1BG) 3. Alpha-1-acid glycoprotein (AGP) 4. Alpha-2 macroglobulin (A2M) 5. Angiotensin II 6. Apolipoprotein D (ApoD) 7. Apolipoprotein H (ApoH) 8. Ceruloplasmin | <ol style="list-style-type: none"> 9. Clusterin 10. Fetuin A 11. Haptoglobin 12. Hemopexin 13. Histidine-proline-rich glycoprotein (HPRG) 14. Immunoglobulin (IgG) 15. Transferrin 16. Vitamin D Binding Protein |
|--|--|

Figure 25. Antibody Array Experimental Set Up. Displayed is the experimental set up used for the glycoprotein antibody array. Each antibody was spotted in duplicate on separate slides, to capture glycoproteins from biological replicates of serum samples.

Utilizing a MALDI-IMS antibody array led to the identification of N-glycans of specific glycoproteins with the capability to distinguish HCC subtypes. As shown in Figure 26, a more sensitive model utilizing N-glycans of specific N-glycans can be identified. Once again, the aim of this model is to differentiate combined S1/S2 from S3 tumors utilizing patient-matching serum. Results demonstrated that the utilization of two N-glycans, $m/z = 1419.479$; M6 from Alpha-1-B Glycoprotein and $m/z = 1765.526$; A2G2 + SO₄ from transferrin, leads to a model with improved sensitivity in differentiating S1/S2 from S3 serum. This model achieved an AUC of 0.881 with a SE of 0.097, and underwent LOOCV with an AUC of 0.802. While the sample size is too small to make sweeping statements regarding clinical utility, the performance of this model emphasizes the ability of differential N-glycosylation of serum glycoproteins to distinguish subtypes of HCC tumors.

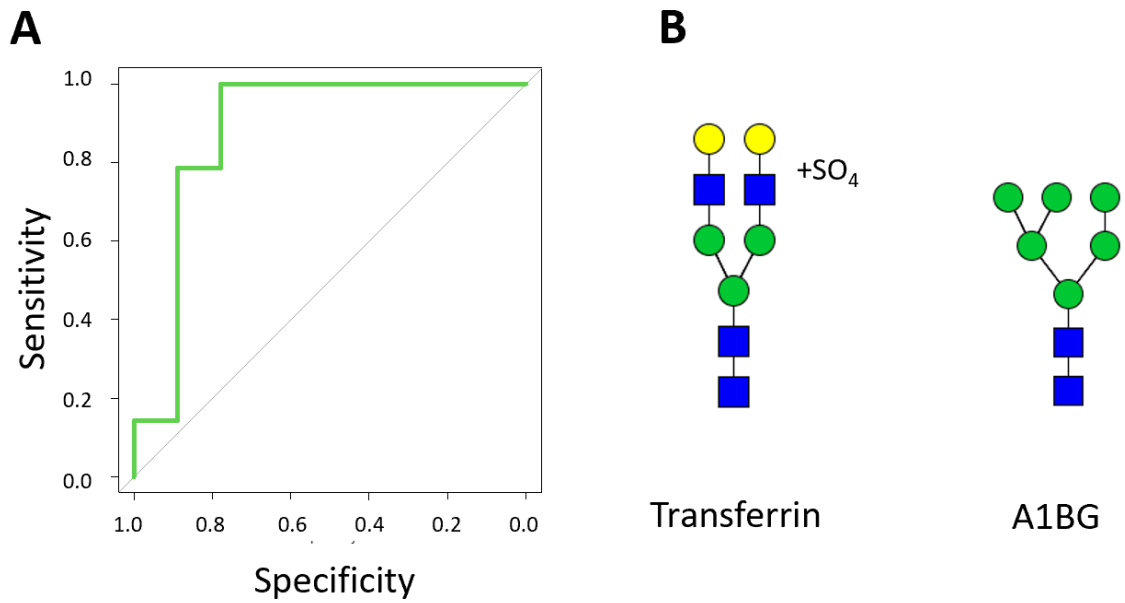


Figure 26. Glycoprotein-Specific Model to Differentiate Subtypes. (A): The ROC curve used to differentiate S1/S2 from S3 serum using antibody array-based N-glycosylation. This model includes an AUC of 0.881. (B): The two N-glycans included within this model are and $m/z = 1765.526$ from transferrin (left), and $m/z = 1419.479$ from A1BG (right).

5.4 Discussion

In recent years, aberrant N-glycosylation has been a major focus of cancer biomarker development, which includes altered N-glycosylation of known biomarkers such as AFP and PSA.^{254,305} It has been clearly demonstrated that N-glycosylation of cancer is widely dysregulated from non-cancer, although much of this work, particularly that done through MALDI-IMS, has focused on tissue N-glycosylation.^{235,270} In order to extend this work to clinically relevant biomarker development, it must be demonstrated that cancer-caused dysregulation of N-glycosylation is apparent in serum as well as tumor tissue. To accomplish this, novel MALDI-IMS techniques have been developed aiming to analyze N-glycosylation of serum glycoproteins and connect with spatial analyses of tumor tissue N-glycosylation.^{287,288,304} In this work, we utilized these serum MALDI-IMS techniques to analyze patient-matching serum and tissue samples in an effort to connect tumor-related tissue N-glycosylation to tumor-related serum N-glycosylation in a subtype-dependent fashion. Additionally, we utilized background liver tissue samples to validate the differentiation of N-glycosylation between HCC and non-HCC tissue. Results demonstrated HCC subtype-related differences in both serum and tissue, although there was less distinction between S1 and S2 tumors than what was previously reported.

In comparing the N-glycosylation of HCC tissue that of background liver tissue, several interesting results emerged. Most importantly, it remained clearly evident that N-glycosylation is differentiable between HCC and non-HCC, with a linear regression model consisting of multiple N-glycans doing so very accurately with an AUC of 0.9842.

However, the three N-glycans utilized in this model were somewhat unexpected, with no fucosylated N-glycans and only one tetra-antennary N-glycan. It's important to note that this sample set is roughly 50% S3 tumors, which have noticeably less cancer-related fucosylation. The sample set skewing towards S3 tumors makes sense, as these are less clinically aggressive tumors that are more likely to be diagnosed in time to resect. However, there are enough S1/S2 tumors included to confidently state that this N-glycan model is capable of distinguishing tumors of all three subtypes from the matching background liver tissues. Although this N-glycan based model is highly accurate, it relies on tissue samples to distinguish cancer from non-cancer, which makes detracts from its clinical relevance, as there are simpler ways based on histology to identify HCC from tissue samples. Regardless, an accurate, N-glycan based model to distinguish HCC tissue from non-HCC tissue is highly valuable in providing additional validation that N-glycosylation is significantly dysregulated within HCC.

Results from subtype-dependent N-glycosylation of tissue offered new insights into the varying N-glycosylation of each subtype, although it remained clear that each subtype has unique enough N-glycosylation to distinguish subtypes based on N-glycans alone. Previous results suggested that S1 had highly dysregulated N-glycosylation, S2 had decreased fucosylation, and S3 was most similar to non-HCC tissue although still distinct based on tetra-antennary glycosylation.¹¹⁵ Results from this sample cohort clearly demonstrated that S1/S2 tumors were highly unique from S3 tumors, with there being 11 N-glycans with statistically significantly different expression between S1 and S3 and 18 N-glycans with statistically significantly different expression between S2 and S3.

The focal point of the difference between both of S1/S2 and S3 tumors appears to be fucosylation, with a large number of fucosylated N-glycans increased in S1/S2 compared to S3 tumors. However, on the level of individual N-glycans, there was not nearly as large of a difference between S1 and S2 as previously indicated. In order to truly assess fucosylation in regard to the distinction between S1 and S2 tumors, a larger sample set may be required, as this sample cohort consisted of 50% S3 tumors. Regardless, it remained possible to identify a model capable of accurately distinguishing all three subtypes with an overall AUC of 0.903 and an accuracy of 0.7586.

In regard to serum N-glycosylation, there was a remarkable amount of correlation between total serum N-glycosylation and tumor-specific N-glycosylation, with 24.7% of individual N-glycans significantly positively correlated between the two. Considering that many of the glycoproteins within serum did not originate from the tumor, or even from the liver, that level of correlation suggests that there is significant HCC tumor biomarker potential contained within the serum N-glycome. This work also analyzed whether there are subtype-dependent differences in serum, which would significantly further the idea that utilizing tumor subtypes is a viable way to improve cancer biomarker development. Due to sample size issues (only 3 S2 serum samples), S1 and S2 serum samples were combined, with an effort to distinguish S1/S2 from S3. With a total serum N-glycan analysis, it was possible to utilize two N-glycans to distinguish these two groupings with an AUC of 0.818. When focusing on specific glycoproteins utilizing an antibody array, the classification ability improved, with an AUC of 0.881. These results suggest that specific N-glycans of serum glycoproteins can be effectively

used to distinguish varying subtypes of HCC. This concept has clear prognostic capabilities, but it also suggests the potential benefit of identifying glycan-based biomarkers of varying tumor subtypes that could work synergistically to improve sensitivity overall.

Chapter 6: Conclusions, Limitations, and Future Studies

6.1 Overall Findings

As the capabilities of N-glycan profiling through MALDI-IMS for both tissue and serum progress, it becomes a valuable tool for the identification and development of much needed glycan-based biomarkers for cancer. The work presented in this dissertation addresses the application of MALDI-IMS N-glycomics to the identification and development of more sensitive HCC biomarkers, with the novel incorporation of genetic information through tumor subtypes. Throughout this work, we validated tumor subtype-dependent differences in N-glycosylation (Chapter 3), developed a novel MALDI-IMS methodology to target de-sialylated N-glycans (Chapter 4), and analyzed the N-glycosylation of patient-matching tissue and serum (Chapter 5). For all of these works, the conclusions, limitations, and areas in need of future studies are highlighted throughout this chapter.

6.2 Analysis of the N-Glycosylation of HCC Tumor Subtypes

6.2.1 Conclusions

Previous work clearly demonstrated that the N-glycosylation of HCC is distinct from non-HCC liver tissue, which is something that could be visualized through MALDI-IMS. However, this was in a heterogeneous manner that lacked the sensitivity needed for application to clinically relevant biomarker development. In order to address this heterogeneity, HCC tumor subtypes were analyzed in this study, and results showed that there are subtype-dependent differences in N-glycosylation which could be utilized to identify more sensitive glycan-based biomarkers. S1 tumors were shown to have

significantly dysregulated N-glycosylation overall, particularly focused on fucosylation, when compared to adjacent tissue. S2 tumors demonstrated significantly less fucosylation than S1 tumors in work done as part of Chapter 3, but this did not carry over to work done as part of Chapter 5. Further analyses with larger sample sizes are needed to further elucidate N-glycan based differences between S1 and S2 tumors. Between all tissue-based and serum-based analyses it was clear that S1 and S2 tumors are distinct from S3 tumors, which have N-glycosylation that is more similar to non-cancer than either S1 or S2. This was expected, as S3 tumors are less clinically aggressive and more well differentiated. Overall, N-glycosylation of all three subtypes was able to be differentiated in tissue, and combined S1/S2 was able to be differentiated in serum. This work provides evidence through a novel methodology that N-glycosylation of tumors is reliant on genetic and molecular features of the tumor, and biomarker discovery should incorporate this information in order to identify more sensitive biomarkers.

6.2.2 Limitations and Future Studies

A limitation of this study is the need to utilize the Hoshida classification system to subtype tumors prior to N-glycomic analyses. This can be time consuming, and it limits the sample size available for analyses. Additionally, tumor subtyping does not result in subtypes that are entirely distinct from each other, as tumors exist on a gradient in regard to genetic/molecular features. In order to further progress this research, N-glycosylation could be correlated with specific genetic or molecular pathway

features, such as through spatial RNA-seq analyses. This would allow for better elucidation of mechanisms, as N-glycan features could be specifically linked to genetic/molecular features of interest as opposed to broadly linked to tumor subtypes. Additionally, such an approach may allow for larger clinical sample sizes, as the need to subtype tumors through the Hoshida classification system would be eliminated. Another future study would be to develop N-glycan based tumor subtypes and to correlate these subtypes with clinical outcomes and treatment responses, along with use to further identify N-glycan-based diagnostic biomarkers. This would be another approach for the application of N-glycomics to cancer detection and treatment, which is the overall goal of this study.

6.3 Matching Serum and Tissue N-Glycan Analysis for HCC Biomarker Development

6.3.1 Conclusions

In order for the dysregulated N-glycosylation of HCC tissue to be relevant to clinical biomarker development, it must be demonstrated that differentiable tumor N-glycosylation translates to the N-glycosylation of serum glycoproteins. Considering that a large percentage of serum glycoproteins originate from the liver, we hypothesized that this would be the case. Results showed a remarkable level of correlation between tumor-specific N-glycosylation and the total N-glycome of serum within a sample set of patient-matching samples. Additionally, it was demonstrated that N-glycan-based models can effectively discriminate between S1/S2 tumors and S3 tumors with both total serum N-glycomic data and glycoprotein-specific N-glycomic data. This validation

of subtype-dependent differences in serum is incredibly important in establishing that genetic differences between tumors lead to not only differences in tissue N-glycosylation but also differences in serum N-glycosylation. Overall, this work provided evidence that tissue N-glycosylation of HCC is highly relevant to the N-glycosylation of serum glycoproteins, in a manner that is encouraging for future N-glycan based biomarker studies.

6.3.2 Limitations and Future Studies

In order to perform this work, a sample cohort of patient-matching tumor tissue and serum was required, which allowed for control over demographic/ethnicity-based differences that could contribute to N-glycosylation differences. However, this resulted in a sample size of only 23 serum samples, which limited the power of statistical analyses that could be performed. In order to further validate the applicability of tissue-based cancer N-glycosylation to glycan-based serum biomarker discovery, a larger sample cohort is needed of either patient-matching tissue and serum samples or samples that control for demographic, ethnicity, and etiology-based differences between the tissue and serum sample sets. Additionally, another study for the future is the comparison of HCC serum to cirrhotic serum samples to further validate cancer-related differences. While this is work that has been previously done in literature, it has not been analyzed in a subtype-dependent manner.

6.4 Utilization of Sialidase for MALDI-IMS N-Glycan Imaging

6.4.1 Conclusions

Chapter 4 focused on the adaptation of the exoglycosidase sialidase into the MALDI-IMS N-glycan imaging workflow. The purpose of this was to increase sensitivity for non-sialylated N-glycans as well as to remove the m/z peaks of sialylated N-glycans that may interfere with the m/z peaks of multi-fucosylated N-glycans. Results showed clear applicability of Sialidase Prime to the MALDI-IMS workflow, with simultaneous PNGase F Prime cleavage of N-glycans from their associated glycoproteins. This workflow was then utilized for both tissue and serum workflows in Chapter 5. Not only was it possible to remove m/z peaks of sialylated N-glycans, the efficiency of enzymatic cleavage of N-glycans was improved, which was observed through the increased signal intensity of both de-sialylated N-glycans and N-glycans that biologically cannot have a sialic acid. This resulted in the development of a novel methodology specifically able to analyze non-sialylated N-glycans with high sensitivity through MALDI-IMS.

6.4.2 Limitations and Future Studies

The obvious limitation of this approach is the loss of sialic acid information, which is often biologically relevant. However, this approach is intended for circumstances in which the N-glycans of interest are either not sialylated or the sialylation is not important to the scientific question being addressed, and in these cases utilizing sialidase within MALDI-IMS N-glycan imaging is a clearly valuable methodology. Future studies in this area should focus on the application of additional

endoglycosidases and exoglycosidases, both individually and synergistically. Thus far, published work has primarily focused on PNGase F, endo F3, and sialidase, but additional enzymes such as mannosidases, galactosidases, sulfatases, and Endoglycosidase H may all provide value within the established MALDI-IMS workflow.

6.5 Final Thoughts

The work in this dissertation has described the novel application of tumor subtype information into N-glycomics-based analyses of HCC tissues and serum for clinical biomarker development. The results of this work have indicated that incorporating genetically based subtypes of tumors within N-glycan analyses yields a better understanding of the N-glycosylation of HCC. While this work has not identified an individual N-glycan to act as a biomarker for a subtype of HCC, which would require much larger sample sizes, it has established that there is tumor subtype-dependent N-glycosylation in both HCC tissue and serum. Additionally, the positive correlation of N-glycan expression between tumor tissue and serum has furthered the idea that it is possible to identify N-glycan-based biomarkers for HCC within serum, which would revolutionize how HCC can be detected clinically. Overall, this dissertation represents yet another scientific area in which MALDI-IMS N-glycan imaging is an incredibly valuable resource.

REFERENCES

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Society, A. C. *Cancer Facts & Figures 2022.* American Cancer Society (2022) doi:10.3238/arztebl.2008.0255.
3. Marrero, J. A. *et al.* Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* **68**, 723–750 (2018).
4. Tsuchiya, N. *et al.* Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J. Gastroenterol.* **21**, 10573–10583 (2015).
5. Yarchoan, M. *et al.* Recent developments and therapeutic strategies against hepatocellular carcinoma. *Cancer Res.* **79**, 4326–4330 (2019).
6. Schuppan, D. & Afdhal, N. H. Liver cirrhosis. *Lancet* **371**, 838–851 (2008).
7. Refolo, M. G., Messa, C., Guerra, V., Carr, B. I. & D'alessandro, R. Inflammatory mechanisms of HCC development. *Cancers (Basel).* **12**, 1–23 (2020).
8. Forner, A., Reig, M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **391**, 1301–1314 (2018).
9. El-Serag, H. B. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology* **142**, 1264–1273 (2012).
10. Singal, A. K. *et al.* Antiviral Therapy Reduces Risk of Hepatocellular Carcinoma in Patients With Hepatitis C Virus-Related Cirrhosis. *Clin. Gastroenterol. Hepatol.* **8**, 192–199 (2010).
11. McMahon, B. J. *et al.* Relationship Between Level of Hepatitis B Virus DNA and Liver Disease: A Population-based Study of Hepatitis B e Antigen-Negative Persons With Hepatitis B. *Clin. Gastroenterol. Hepatol.* **12**, 701–706.e3 (2014).
12. Shi, J., Zhu, L., Liu, S. & Xie, W. F. A meta-analysis of case-control studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma in China. *Br. J. Cancer* **92**, 607–612 (2005).
13. Choi, D. T. *et al.* Hepatocellular Carcinoma Screening Is Associated With Increased Survival of Patients With Cirrhosis. *Clin. Gastroenterol. Hepatol.* **17**, 976–987 (2019).
14. Llovet, J. M., Brú, C. & Bruix, J. Prognosis of hepatocellular carcinoma: The BCLC staging classification. *Semin. Liver Dis.* **19**, 329–337 (1999).
15. Forner, A., Reig, M. E., Rodriguez De Lope, C. & Bruix, J. Current strategy for staging and treatment: The BCLC update and future prospects. *Seminars in Liver Disease* vol. 30 61–74 (2010).
16. El-Serag, H. B. & Kanwal, F. Epidemiology of hepatocellular carcinoma in the United States: Where are we? Where do we go? *Hepatology* **60**, 1767–1775 (2014).
17. Chen, C. *et al.* Risk of Hepatocellular Carcinoma Across a Biological Gradient of Serum Hepatitis B Virus DNA Level. *JAMA* **295**, 65–73 (2006).
18. Chan, H. L. Y., Hussain, M. & Lok, A. S. F. Different hepatitis B virus genotypes are

- associated with different mutations in the core promoter and precore regions during hepatitis B e antigen seroconversion. *Hepatology* **29**, 976–984 (1999).
19. Yang, H. I. *et al.* Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma. *J. Natl. Cancer Inst.* **100**, 1134–1143 (2008).
 20. Ni, Y. H. *et al.* Clinical relevance of hepatitis B virus genotype in children with chronic infection and hepatocellular carcinoma. *Gastroenterology* **127**, 1733–1738 (2004).
 21. Kao, J. H., Chen, P. J., Lai, M. Y. & Chen, D. S. Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. *Gastroenterology* **118**, 554–559 (2000).
 22. Chen, J. De *et al.* Carriers of Inactive Hepatitis B Virus Are Still at Risk for Hepatocellular Carcinoma and Liver-Related Death. *Gastroenterology* **138**, 1747–1754.e1 (2010).
 23. Haydon, G. H., Jarvis, L. M., Simmonds, P. & Hayes, P. C. Association between chronic hepatitis C infection and hepatocellular carcinoma. *The Lancet* vol. 345 928–929 (1995).
 24. Kirk, G. D. *et al.* HIV, age, and the severity of hepatitis C virus-related liver disease: A cohort study. *Ann. Intern. Med.* **158**, 658–667 (2013).
 25. Raimondi, S., Bruno, S., Mondelli, M. U. & Maisonneuve, P. Hepatitis C virus genotype 1b as a risk factor for hepatocellular carcinoma development: A meta-analysis. *J. Hepatol.* **50**, 1142–1154 (2009).
 26. Li, Z., Tuteja, G., Schug, J. & Kaestner, K. H. Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer. *Cell* **148**, 72–83 (2012).
 27. Yu, M. W. *et al.* Androgen receptor exon 1 CAG repeat length and risk of hepatocellular carcinoma in women. *Hepatology* **36**, 156–163 (2002).
 28. Donato, F. *et al.* Alcohol and hepatocellular carcinoma: The effect of lifetime intake and hepatitis virus infections in men and women. *Am. J. Epidemiol.* **155**, 323–331 (2002).
 29. Morgan, T. R., Mandayam, S. & Jamal, M. M. Alcohol and hepatocellular carcinoma. in *Gastroenterology* vol. 127 87–96 (2004).
 30. Dyson, J. *et al.* Hepatocellular cancer: The impact of obesity, type 2 diabetes and a multidisciplinary team. *J. Hepatol.* **60**, 110–117 (2014).
 31. Karagozian, R., Derdák, Z. & Baffy, G. Obesity-associated mechanisms of hepatocarcinogenesis. *Metabolism* **63**, 607–617 (2014).
 32. Sanyal, A., Poklepovic, A., Moyneur, E. & Barghout, V. Population-based risk factors and resource utilization for HCC: US perspective. *Curr. Med. Res. Opin.* **26**, 2183–2191 (2010).
 33. Kanwal, F. *et al.* Trends in the Burden of Nonalcoholic Fatty Liver Disease in a United States Cohort of Veterans. *Clin. Gastroenterol. Hepatol.* **14**, 301–308.e2 (2016).
 34. Schlesinger, S. *et al.* Diabetes mellitus, insulin treatment, diabetes duration, and risk of biliary tract cancer and hepatocellular carcinoma in a European Cohort. *Ann. Oncol.* **24**, 2449–2455 (2013).
 35. Nordenstedt, H., White, D. L. & El-Serag, H. B. The changing pattern of

- epidemiology in hepatocellular carcinoma. *Dig. Liver Dis.* **42**, S206–S214 (2010).
36. Ho, D. W. H., Lo, R. C. L., Chan, L. K. & Ng, I. O. L. Molecular pathogenesis of hepatocellular carcinoma. *Liver Cancer* vol. 5 290–302 (2016).
 37. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
 38. Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014).
 39. Schulze, K., Nault, J. C. & Villanueva, A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J. Hepatol.* **65**, 1031–1042 (2016).
 40. Wheeler, D. A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327–1341.e23 (2017).
 41. Hoshida, Y. *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).
 42. Hoshida, Y. *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* **69**, 7385–7392 (2009).
 43. Forner, A. *et al.* Diagnosis of Hepatic Nodules 20 mm or Smaller in Cirrhosis: Prospective Validation of the Noninvasive Diagnostic Criteria for Hepatocellular Carcinoma. *Hepatology* **47**, 97–104 (2008).
 44. DelaCourt, A. T. & Mehta, A. S. Liver Cancer (Current Therapies). in *Reference Module in Biomedical Sciences* 1–14 (Elsevier, 2021). doi:10.1016/b978-0-12-820472-6.00007-4.
 45. Dufour, J. F. *et al.* EASL – EORTC Clinical Practice Guidelines : Management of hepatocellular carcinoma. *J. Hepatol.* **56**, 908–943 (2012).
 46. Ishizawa, T. *et al.* Neither Multiple Tumors Nor Portal Hypertension Are Surgical Contraindications for Hepatocellular Carcinoma. *Gastroenterology* **134**, 1908–1916 (2008).
 47. Mazzaferro, V., Regalia, E., Doci, R., Andreola, S. & Pulvirenti, A. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N. Engl. J. Med.* **334**, 693–699 (1996).
 48. Cucchetti, A. *et al.* Cost-effectiveness of hepatic resection versus percutaneous radiofrequency ablation for early hepatocellular carcinoma. *J. Hepatol.* **59**, 300–307 (2013).
 49. Hasegawa, K. *et al.* Comparison of resection and ablation for hepatocellular carcinoma : A cohort study based on a Japanese nationwide survey. *J. Hepatol.* **58**, 724–729 (2013).
 50. Llovet, J. M. & Bruix, J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. *Hepatology* **37**, 429–442 (2003).
 51. Burrel, M. *et al.* Survival of patients with hepatocellular carcinoma treated by transarterial chemoembolisation (TACE) using Drug Eluting Beads . Implications for clinical practice and trial design. *J. Hepatol.* **56**, 1330–1335 (2012).
 52. Yoon, S. M. *et al.* Efficacy and Safety of Transarterial Chemoembolization Plus

- External Beam Radiotherapy vs Sorafenib in Hepatocellular Carcinoma With Macroscopic Vascular Invasion A Randomized Clinical Trial. *JAMA Oncol.* **4**, 661–669 (2018).
53. Kudo, M. *et al.* Brivanib as Adjuvant Therapy to Transarterial Chemoembolization in Patients With Hepatocellular Carcinoma: A Randomized Phase III Trial. *Hepatology* **60**, 1697–1707 (2014).
 54. Lencioni, R. *et al.* Sorafenib or placebo plus TACE with doxorubicin-eluting beads for intermediate stage HCC : The SPACE trial. *J. Hepatol.* **64**, 1090–1098 (2016).
 55. Llovet, J. M., Ricci, S., Mazzaferro, V., Hilgard, P. & Gane, E. Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 2498 (2008).
 56. Bruix, J. *et al.* Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **389**, 56–66 (2017).
 57. Wilhelm, S. M. *et al.* BAY 43-9006 Exhibits Broad Spectrum Oral Antitumor Activity and Targets the RAF/MEK/ERK Pathway and Receptor Tyrosine Kinases Involved in Tumor Progression and Angiogenesis. *Cancer Res.* **64**, 7099–7109 (2004).
 58. Wilhelm, S. M. *et al.* Regorafenib (BAY 73-4506): A new oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases with potent preclinical antitumor activity. *Int. J. Cancer* **129**, 245–255 (2011).
 59. Cheng, A. *et al.* Sunitinib Versus Sorafenib in Advanced Hepatocellular Cancer : Results of a Randomized Phase III Trial. *J. Clin. Oncol.* **31**, 4067–4076 (2013).
 60. Johnson, P. J. *et al.* Brivanib Versus Sorafenib As First-Line Therapy in Patients With Unresectable , Advanced Hepatocellular Carcinoma : Results From the Randomized Phase III BRISK-FL Study. *J. Clin. Oncol.* **31**, 3517–3524 (2013).
 61. Sherman, M., Bch, M. B. & Frcp, C. Surveillance for hepatocellular carcinoma. *Best Pract. Res. Clin. Gastroenterol.* **28**, 783–793 (2014).
 62. Tzartzeva, K. *et al.* Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta- analysis. *Gastroenterology* **154**, 1706–1718 (2018).
 63. Simmons, O. *et al.* Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Aliment Pharmacol Ther.* **45**, 169–177 (2017).
 64. Farvardin, S. *et al.* Patient-reported barriers are associated with lower hepatocellular carcinoma surveillance rates in patients with cirrhosis. *Hepatology* **65**, 875–884 (2017).
 65. Lok, A. S. *et al.* Des-gamma-carboxy Prothrombin and Alpha fetoprotein as Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Gastroenterology* **138**, 1–21 (2010).
 66. Chang, T. *et al.* Alpha-Fetoprotein Measurement Benefits Hepatocellular Carcinoma Surveillance in Patients with Cirrhosis. *Am. J. Gastroenterol.* **110**, 836–844 (2015).
 67. Parikh, N. D. *et al.* Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **29**, 2495–2503 (2020).

68. Lee, E. *et al.* Improving Screening for Hepatocellular Carcinoma by Incorporating Data on Levels of alpha-Fetoprotein , Over Time. *Clin. Gastroenterol. Hepatol.* **11**, 437–440 (2013).
69. Taketa, K. *et al.* A Collaborative Study for the Evaluation of Lectin-Reactive a-Fetoproteins in Early Detection of Hepatocellular Carcinoma. *Cancer Res.* **53**, 5419–5424 (1993).
70. Li, D., Mallory, T. & Satomura, S. AFP-L3 : a new generation of tumor marker for hepatocellular carcinoma. *Clin. Chim. Acta* **313**, 15–19 (2001).
71. Choi, J. *et al.* Longitudinal Assessment of Three Serum Biomarkers to Detect Very Early-Stage Hepatocellular Carcinoma. *Hepatology* **69**, 1983–1994 (2019).
72. Sezaki, H., Akuta, N., Suzuki, F., Suzuki, Y. & Saitoh, S. Highly sensitive AFP-L3 assay is useful for predicting recurrence of hepatocellular carcinoma after curative treatment pre- and postoperatively. *Hepatol. Res.* **41**, 1036–1045 (2011).
73. Tsai, S.-L. S.-L. *et al.* Plasma Des-γ-Carboxyprothrombin in the Early Stage of Hepatocellular Carcinoma. *Hepatology* **11**, 481–487 (1989).
74. Nakamura, S. *et al.* Sensitivity and Specificity of Des-Gamma-Carboxy Prothrombin for Diagnosis of Patients with Hepatocellular Carcinomas Varies According to Tumor Size. *Am. J. Gastroenterol.* **101**, 2038–2043 (2006).
75. Wan, H. *et al.* Comparison osteopontin vs AFP for the diagnosis of HCC : A meta-analysis. *Clin. Res. Hepatol. Gastroenterol.* **38**, 706–714 (2014).
76. Sun, T., Tang, Y., Diwen, S., Bu, Q. & Li, P. Osteopontin versus alpha-fetoprotein as a diagnostic marker for hepatocellular carcinoma: a meta-analysis. *Oncotargets Ther.* **11**, 8925–8935 (2018).
77. Zhu, W.-W., Guo, J.-J., Guo, L., Jia, H.-L. & Ye, Q.-H. Evaluation of Midkine as a Diagnostic Serum Biomarker in Hepatocellular Carcinoma. *Clin. Cancer Res.* **19**, 3944–3954 (2013).
78. Vongsuvan, R. *et al.* Midkine increases diagnostic yield in AFP negative and NASH-related hepatocellular carcinoma. *PLoS One* **11**, 1–12 (2016).
79. Block, T. M. *et al.* Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 779–784 (2005).
80. Marrero, J. A. *et al.* GP73 , a resident Golgi glycoprotein , is a novel serum marker for hepatocellular carcinoma. *Hepatology* **43**, 1007–1012 (2005).
81. Shen, Q. *et al.* Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. *Lancet Oncol.* **13**, 817–826 (1998).
82. Ishizuka, H., Nakayama, T., Matsuoka, S., Gotoh, I. & Ogawa, M. Prediction of the Development of Hepato-Cellular-Carcinoma in Patients with Liver Cirrhosis by the Serial Determinations of Serum Alpha-L-fucosidase Activity. *Intern. Med.* **38**, 927–931 (1999).
83. Giannelli, G. *et al.* SCCA antigen combined with alpha-fetoprotein as serologic markers of HCC. *Int. J. Cancer* **117**, 506–509 (2005).
84. Jia, X., Liu, J., Gao, Y., Huang, Y. & Du, Z. Diagnosis Accuracy of Serum Glypican-3 in Patients with Hepatocellular Carcinoma: A Systematic Review with Meta-

- analysis. *Arch. Med. Res.* **45**, 580–588 (2014).
85. Johnson, P. J. *et al.* The detection of hepatocellular carcinoma using a prospectively developed and validated model based on serological biomarkers. *Cancer Epidemiol. Biomarkers Prev.* **23**, 144–153 (2014).
 86. Berhane, S. *et al.* Role of the GALAD and BALAD-2 Serologic Models in Diagnosis of Hepatocellular Carcinoma and Prediction of Survival in Patients. *Clin. Gastroenterol. Hepatol.* **14**, 875–886 (2016).
 87. Wang, M. *et al.* The Doylestown algorithm – a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prev. Res.* **9**, 172–179 (2017).
 88. Mehta, A. S. *et al.* Application of the Doylestown algorithm for the early detection of hepatocellular carcinoma. *PLoS One* **13**, 1–12 (2018).
 89. Wang, M. *et al.* Changes in the glycosylation of kininogen and the development of a kininogen based algorithm for the early detection of HCC. *Cancer Epidemiol. Biomarkers Prev.* **26**, 795–803 (2017).
 90. Nguema-Ona, E. *et al.* Cell wall O-glycoproteins and N-glycoproteins: Aspects of biosynthesis and function. *Front. Plant Sci.* **5**, (2014).
 91. Kudelka, M. R., Ju, T., Heimburg-Molinaro, J. & Cummings, R. D. Simple sugars to complex disease-mucin-type O-glycans in cancer. in *Advances in Cancer Research* vol. 126 53–135 (Academic Press Inc., 2015).
 92. Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Mol. Cell. Proteomics* **12**, 846–855 (2013).
 93. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**, 346–366 (2019).
 94. Stanley, P., Taniguchi, N. & Aebi, M. *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press (Cold Spring Harbor Laboratory Press, 2017). doi:10.1101/GLYCOBIOLOGY.3E.009.
 95. Bennett, E. P. *et al.* Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736–756 (2012).
 96. Wenekes, T. *et al.* Glycosphingolipids - Nature, function, and pharmacological modulation. *Angewandte Chemie - International Edition* vol. 48 8848–8869 (2009).
 97. Iozzo, R. V & Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biology* vol. 42 11–55 (2015).
 98. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
 99. Okamoto, R., Izumi, M. & Kajihara, Y. Decoration of proteins with sugar chains: Recent advances in glycoprotein synthesis. *Curr. Opin. Chem. Biol.* **22**, 92–99 (2014).
 100. Johansson, M. E. V., Sjövall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 352–361 (2013).
 101. Lingwood, C. A. & Hakomori, S. Selective inhibition of cell growth and associated changes in glycolipid metabolism induced by monovalent antibodies to

- glycolipids. *Exp. Cell Res.* **108**, 385–391 (1977).
102. Weerapana, E. & Imperiali, B. Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91–101 (2006).
 103. Loomes, K. M., Senior, H. E., West, P. M. & Robertson, A. M. Functional protective role for mucin glycosylated repetitive domains. *Eur. J. Biochem.* **266**, 105–111 (1999).
 104. Itin, C., Roche, A. C., Monsigny, M. & Hauri, H. P. ERGIC-53 is a functional mannose-selective and calcium-dependent human homologue of leguminous lectins. *Mol. Biol. Cell* **7**, 483–493 (1996).
 105. Marquardt, T. *et al.* Correction of Leukocyte Adhesion Deficiency Type II With Oral Fucose. *Blood* **94**, 3976–3985 (1999).
 106. Martinon, F., Mayor, A. & Tschopp, J. The inflammasomes: Guardians of the body. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
 107. Esko, J. D. & Sharon, N. Microbial Lectins: Hemagglutinins, Adhesins, and Toxins. in *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009).
 108. Vaishnava, S. *et al.* The Antibacterial Lectin RegIII. *Science (80-.)*. **334**, 255–258 (2011).
 109. Freire-De-Lima, L., Fonseca, L. M., Oeltmann, T., Mendonça-Previato, L. & Previato, J. O. The trans-sialidase, the major *Trypanosoma cruzi* virulence factor: Three decades of studies. *Glycobiology* **25**, 1142–1149 (2015).
 110. Khatua, B., Bhattacharya, K. & Mandal, C. Sialoglycoproteins adsorbed by *Pseudomonas aeruginosa* facilitate their survival by impeding neutrophil extracellular trap through siglec-9. *J. Leukoc. Biol.* **91**, 641–655 (2012).
 111. Spiro, R. G. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, (2002).
 112. Herscovics, A. Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochimica et Biophysica Acta - General Subjects* vol. 1473 96–107 (1999).
 113. Berninsone, P. M. & Hirschberg, C. B. Nucleotide sugar transporters of the Golgi apparatus. *Current Opinion in Structural Biology* vol. 10 542–547 (2000).
 114. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
 115. DelaCourt, A. *et al.* N-glycosylation patterns correlate with hepatocellular carcinoma genetic subtypes. *Mol. Cancer Res.* **19**, 1868–1877 (2021).
 116. Pinho, S. S. & Reis, C. A. Glycosylation in cancer : mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
 117. Gill, D. J., Chia, J., Senewiratne, J. & Bard, F. Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *J. Cell Biol.* **189**, 843–858 (2010).
 118. Kellokumpu, S., Sormunen, R. & Kellokumpu, I. Abnormal glycosylation and altered Golgi structure in colorectal cancer: Dependence on intra-Golgi pH. *FEBS Lett.* **516**, 217–224 (2002).
 119. Keeley, T. S., Yang, S. & Lau, E. The diverse contributions of fucose linkages in cancer. *Cancers (Basel)*. **11**, 1241 (2019).
 120. Christiansen, M. N. *et al.* Cell surface protein glycosylation in cancer. *Proteomics*

- 14**, 525–546 (2014).
121. Hakomori, S. Glycosylation defining cancer malignancy: New wine in an old bottle. *Proceedings of the National Academy of Sciences of the United States of America* vol. 99 10231–10233 (2002).
 122. Arnold, J. N., Saldova, R., Hamid, U. M. A. & Rudd, P. M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**, 3284–3293 (2008).
 123. Renkonen, J., Paavonen, T. & Renkonen, R. Endothelial and epithelial expression of sialyl Lewis and sialyl Lewis in lesions of breast carcinoma. *Int. J. Cancer* **74**, 296–300 (1997).
 124. Ugorski, M. & Laskowska, A. Sialyl Lewis a : a tumor-associated carbohydrate antigen in-volved in adhesion and metastatic potential of cancer cells *. *Acta Biochim. Pol.* **49**, 303–311 (2002).
 125. Davidson, B. *et al.* Expression of carbohydrate antigens in advanced-stage ovarian carcinomas and their metastases - A clinicopathologic study. *Gynecol. Oncol.* **77**, 35–43 (2000).
 126. Kim, Y. J. & Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate Journal* vol. 14 569–576 (1997).
 127. McDowell, C. T. *et al.* Imaging Mass Spectrometry and Lectin Analysis of N-linked Glycans in Carbohydrate Antigen Defined Pancreatic Cancer Tissues. *Mol. Cell. Proteomics* **20**, 100012 (2021).
 128. Hebbler, M., Krzewinski-Recchi, M.-A., Hornez, L. & Verdiere, A. Prognostic value of tumoral sialyltransferase expression and circulating E-selectin concentrations in node-negative breast cancer patients. *Int. J. Biol. Markers* **18**, 116–122 (2003).
 129. Christie, D. R., Shaikh, F. M., Lucas, J. A., Lucas, J. A. & Bellis, S. L. ST6Gal-I expression in ovarian cancer cells promotes an invasive phenotype by altering integrin glycosylation and function. *J. Ovarian Res.* **1**, 1–8 (2008).
 130. Wi, G. R. *et al.* A lectin-based approach to detecting carcinogenesis in breast tissue. *Oncol. Lett.* **11**, 3889–3895 (2016).
 131. Osumi, D. *et al.* Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer Sci.* **100**, 888–895 (2009).
 132. Watanabe, K. *et al.* Fucosylation is associated with the malignant transformation of intraductal papillary mucinous neoplasms: a lectin microarray-based study. *Surg. Today* **46**, 1217–1223 (2016).
 133. Honma, R. *et al.* Expression of fucosyltransferase 8 is associated with an unfavorable clinical outcome in non-small cell lung cancers. *Oncology* **88**, 298–308 (2015).
 134. Kim, H. J. *et al.* Aberrant sialylation and fucosylation of intracellular proteins in cervical tissue are critical markers of cervical carcinogenesis. *Oncol. Rep.* **31**, 1417–1422 (2014).
 135. Takahashi, M., Kuroki, Y., Ohtsubo, K. & Taniguchi, N. Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydrate Research* vol. 344 1387–1390 (2009).
 136. Comunale, M. A. *et al.* Linkage specific fucosylation of alpha-1-antitrypsin in liver

- cirrhosis and cancer patients: Implications for a biomarker of hepatocellular carcinoma. *PLoS One* **5**, 1–9 (2010).
137. Norton, P. A. & Mehta, A. S. Expression of genes that control core fucosylation in hepatocellular carcinoma: Systematic review. *World J. Gastroenterol.* **25**, 2947–2960 (2019).
 138. Miyoshi, E. *et al.* Fucosylation is a promising target for cancer diagnosis and therapy. *Biomolecules* vol. 2 34–45 (2012).
 139. Blomme, B., Van Steenkiste, C., Callewaert, N. & Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *J. Hepatol.* **50**, 592–603 (2009).
 140. Potapenko, I. O. *et al.* Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Molecular Oncology* vol. 4 98–118 (2010).
 141. Taniguchi, N. & Kizuka, Y. Glycans and cancer: Role of N-Glycans in cancer biomarker, progression and metastasis, and therapeutics. in *Advances in Cancer Research* vol. 126 11–51 (Academic Press Inc., 2015).
 142. Wang, Y. *et al.* Loss of a 1, 6-fucosyltransferase inhibits chemical-induced hepatocellular carcinoma and tumorigenesis by down-regulating several cell signaling pathways. *FASEB J.* **29**, 3217–3227 (2015).
 143. Chen, C. Y. *et al.* Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 630–635 (2013).
 144. Lise, M. *et al.* Clinical Correlations of α 2,6-Sialyltransferase Expression in Colorectal Cancer Patients. *Hybridoma* **19**, 281–286 (2000).
 145. Sewell, R. *et al.* The ST6GalNAC-I sialyltransferase localizes throughout the golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J. Biol. Chem.* **281**, 3586–3594 (2006).
 146. Rillahan, C. D. *et al.* Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat. Chem. Biol.* **8**, 661–668 (2012).
 147. Granovsky, M. *et al.* Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nat. Med.* **6**, 306–312 (2000).
 148. Taniguchi, N. & Korekane, H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Reports* vol. 44 772–781 (2011).
 149. Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M. L. & Kerbel, R. S. Beta 1-6 Branching of Asn-Linked Oligosaccharides Is Directly Associated with Metastasis. *Science (80-.)*. **236**, 582–585 (1986).
 150. Lauc, G., Pezer, M., Rudan, I. & Campbell, H. Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta - Gen. Subj.* **1860**, 1574–1582 (2016).
 151. Ferreira, I. G. *et al.* Glycosylation as a main regulator of growth and death factor receptors signaling. *Int. J. Mol. Sci.* **19**, 1–28 (2018).
 152. Lau, K. S. *et al.* Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell* **129**, 123–134 (2007).
 153. Takahashi, M., Tsuda, T., Ikeda, Y., Honke, K. & Taniguchi, N. Role of N-glycans in growth factor signaling. *Glycoconj. J.* **20**, 207–212 (2003).
 154. Seidenfaden, R., Krauter, A., Schertzinger, F., Gerardy-Schahn, R. & Hildebrandt,

- H. Polysialic Acid Directs Tumor Cell Growth by Controlling Heterophilic Neural Cell Adhesion Molecule Interactions. *Mol. Cell. Biol.* **23**, 5908–5918 (2003).
155. Lin, S., Kemmner, W., Grigull, S. & Schlag, P. M. Cell surface $\alpha 2, 6$ -sialylation affects adhesion of breast carcinoma cells. *Exp. Cell Res.* **276**, 101–110 (2002).
 156. Ferrer, C. M. *et al.* O-GlcNAcylation Regulates Cancer Metabolism and Survival Stress Signaling via Regulation of the HIF-1 Pathway. *Mol. Cell* **54**, 820–831 (2014).
 157. Gornik, O., Pavić, T. & Lauc, G. Alternative glycosylation modulates function of IgG and other proteins - Implications on evolution and disease. *Biochimica et Biophysica Acta - General Subjects* vol. 1820 1318–1326 (2012).
 158. Liu, F. T. & Rabinovich, G. A. Galectins as modulators of tumour progression. *Nature Reviews Cancer* vol. 5 29–41 (2005).
 159. Ornstein, D. K. & Pruthi, R. S. Prostate-specific antigen. *Expert Opin. Pharmacother.* **1**, 1399–1411 (2000).
 160. Dow, M. *et al.* Integrative genomic analysis of mouse and human hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9879–E9888 (2018).
 161. Barry, M. J. Screening for Prostate Cancer — The Controversy That Refuses to Die. *N. Engl. J. Med.* **360**, 1351–1354 (2009).
 162. Drake, R. R., Jones, E. E., Powers, T. W. & Nyalwidhe, J. O. Altered glycosylation in prostate cancer. in *Advances in Cancer Research* vol. 126 345–382 (Academic Press Inc., 2015).
 163. Fukushima, K., Satoh, T., Baba, S. & Yamashita, K. $\alpha 1,2$ -Fucosylated and β -N-acetylgalactosaminylated prostate-specific antigen as an efficient marker of prostatic cancer. *Glycobiology* **20**, 452–460 (2009).
 164. Saldoval, R., Fan, Y., Fitzpatrick, J. M., Watson, R. W. G. & Rudd, P. M. Core fucosylation and $\alpha 2$ -3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **21**, 195–205 (2011).
 165. Lang, R. *et al.* Investigation on core-fucosylated prostate-specific antigen as a refined biomarker for differentiation of benign prostate hyperplasia and prostate cancer of different aggressiveness. *Tumor Biol.* **41**, (2019).
 166. Frenette, C. T., Isaacson, A. J., Bargellini, I., Saab, S. & Singal, A. G. A Practical Guideline for Hepatocellular Carcinoma Screening in Patients at Risk. *Mayo Clin. Proc. Innov. Qual. Outcomes* **3**, 302–310 (2019).
 167. Singal, A. G. *et al.* Effectiveness of hepatocellular carcinoma surveillance in patients with cirrhosis. *Cancer Epidemiol. Biomarkers Prev.* **21**, 793–799 (2012).
 168. Marrero, J. A. *et al.* α -Fetoprotein, Des- γ Carboxyprothrombin, and Lectin-Bound α -Fetoprotein in Early Hepatocellular Carcinoma. *Gastroenterology* **137**, 110–118 (2009).
 169. Bast, R. C., Klug, T. L., St. John, E. & Jenison, E. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N. Engl. J. Med.* **309**, 883–887 (1983).
 170. Marcus, C. S., Maxwell, G. L., Darcy, K. M., Hamilton, C. A. & McGuire, W. P. Current approaches and challenges in managing and monitoring treatment

- response in ovarian cancer. *Journal of Cancer* vol. 5 25–30 (2014).
171. Gadducci, A. Serum CA125 assay in the early diagnosis of recurrent epithelial ovarian cancer: A CTF study. *Oncol. Rep.* **3**, 301–303 (1996).
 172. Song, M. J. *et al.* Diagnostic value of CA125 as a predictor of recurrence in advanced ovarian cancer. *Eur. J. Gynaecol. Oncol.* **34**, 148–151 (2013).
 173. Kirwan, A., Utratna, M., O’Dwyer, M. E., Joshi, L. & Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed Research International* vol. 2015 (2015).
 174. Nath, S. & Mukherjee, P. MUC1: A multifaceted oncoprotein with a key role in cancer progression. *Trends in Molecular Medicine* vol. 20 332–342 (2014).
 175. Baldus, S. E., Engelmann, K. & Hanisch, F. G. MUC1 and the MUCs: A family of human mucins with impact in cancer biology. *Critical Reviews in Clinical Laboratory Sciences* vol. 41 189–231 (2004).
 176. Duffy, M. J. Clinical Uses of Tumor Markers: A Critical Review. *Crit. Rev. Clin. Lab. Sci.* **38**, 225–262 (2001).
 177. Chen, W. *et al.* Muc1: Structure, function, and clinic application in epithelial cancers. *Int. J. Mol. Sci.* **22**, 1–15 (2021).
 178. Hsu, C. C. *et al.* Elevated CA19-9 is Associated with Increased Mortality in A Prospective Cohort of Hepatocellular Carcinoma Patients. *Clin. Transl. Gastroenterol.* **6**, (2015).
 179. O’Brien, D. P. *et al.* Serum CA19-9 is significantly upregulated up to 2 years before diagnosis with pancreatic cancer: Implications for early disease detection. *Clin. Cancer Res.* **21**, 622–631 (2015).
 180. Goonetilleke, K. S. & Siriwardena, A. K. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *European Journal of Surgical Oncology* vol. 33 266–270 (2007).
 181. Safi, F., Schlosser, W., Kolb, G. & Beger, H. G. Diagnostic Value of CA 19-9 in Patients with Pancreatic Cancer and Nonspecific Gastrointestinal Symptoms. *J. Gastrointest. Surg.* **1**, 106–112 (1997).
 182. Tang, H. *et al.* Glycans Related to the CA19-9 Antigen Are Increased in Distinct Subsets of Pancreatic Cancers and Improve Diagnostic Accuracy Over CA19-9. *CMGH* **2**, 210-221.e15 (2016).
 183. Liang, Y. *et al.* Differentially expressed glycosylated patterns of α -1-antitrypsin as serum biomarkers for the diagnosis of lung cancer. *Glycobiology* **25**, 331–340 (2014).
 184. Wang, M. *et al.* Biomarker analysis of fucosylated kininogen through depletion of lectin reactive heterophilic antibodies in hepatocellular carcinoma. *J Immunol Methods* **462**, 59–64 (2018).
 185. Singal, A. G. *et al.* Doylestown Plus and GALAD Demonstrate High Sensitivity for HCC Detection in Patients With Cirrhosis. *Clin. Gastroenterol. Hepatol.* **20**, 953-955.e2 (2022).
 186. Dochez, V. *et al.* Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *Journal of Ovarian Research* vol. 12 (2019).
 187. Moore, R. G. *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for

- the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol. Oncol.* **112**, 40–46 (2009).
188. Menon, U. *et al.* Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. *J. Clin. Oncol.* **33**, 2062–2071 (2015).
 189. Yip-Schneider, M. T. *et al.* Biomarker Risk Score Algorithm and Preoperative Stratification of Patients with Pancreatic Cystic Lesions. *J. Am. Coll. Surg.* **233**, 426-434.e4 (2021).
 190. Schwarz, R. T. & Datema, R. Inhibition of the Dolichol Pathway of Protein Glycosylation. *Methods Enzymol.* **83**, 432–443 (1982).
 191. Almahayni, K. *et al.* Small molecule inhibitors of mammalian glycosylation. *Matrix Biol. Plus* 100108 (2022) doi:10.1016/j.mbplus.2022.100108.
 192. Tulsiani, D. R. P., Harris, T. M. & Touster, O. Swainsonine inhibits the biosynthesis of complex glycoproteins by inhibition of Golgi mannosidase II. *J. Biol. Chem.* **257**, 7936–7939 (1982).
 193. Hohenschutz, L. D. *et al.* Castanospermine, A 1,6,7,8-tetrahydroxyoctahydroindolizine alkaloid, from seeds of *Castanospermum australe*. *Phytochemistry* **20**, 811–814 (1981).
 194. Iwami, M. *et al.* A New Immunomodulator, FR-9000494: Taxonomy, Fermentation, Isolation, and Physico-chemical and Biological Characteristics. *J. Antibiot. (Tokyo)*. **40**, 612–622 (1987).
 195. Pijnenborg, J. F. A. *et al.* Cellular Fucosylation Inhibitors Based on Fluorinated Fucose-1-phosphates**. *Chem. - A Eur. J.* **27**, 4022–4027 (2021).
 196. Müller, B., Schaub, C. & Schmidt, R. R. Efficient sialyltransferase inhibitors based on transition-state analogues of the sialyl donor. *Angew. Chemie - Int. Ed.* **37**, 2893–2897 (1998).
 197. Murray, B. W., Wittmann, V., Burkart, M. D., Hung, S. C. & Wong, C. H. Mechanism of human α -1,3-fucosyltransferase V: Glycosidic cleavage occurs prior to nucleophilic attack. *Biochemistry* **36**, 823–831 (1997).
 198. Cummings, R. D. *et al.* Chapter 48: Glycan-Recognizing Probes as Tools. in *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2022). doi:10.1101/GLYCOBIOLOGY.4E.48.
 199. Merkle, R. K. & Cummings, R. D. Lectin Affinity Chromatography of Glycopeptides. *Methods Enzymol.* **138**, 232–259 (1987).
 200. Drake, R. R. *et al.* Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* **5**, 1957–1967 (2006).
 201. Cummings, R. D. Use of lectins in analysis of glycoconjugates. *Methods Enzymol.* **230**, 66–86 (1994).
 202. Bojar, D. *et al.* A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities. *ACS Chem. Biol.* **17**, 2993–3012 (2021).
 203. Allen, A. K. Purification and characterization of an N-acetyllactosamine-specific

- lectin from tubers of *Arum maculatum*. *BBA - Gen. Subj.* **1244**, 129–132 (1995).
204. Sumner, J. B. & Howell, S. F. Identification of Hemagglutinin of Jack Bean with Concanavalin A. *J. Bacteriol.* **32**, 227–237 (1936).
 205. Kaneda, Y. *et al.* The high specificities of Phaseolus vulgaris erythro- and leukoagglutinating lectins for bisecting GlcNAc or β 1-6-linked branch structures, respectively, are attributable to loop B. *J. Biol. Chem.* **277**, 16928–16935 (2002).
 206. Peumans, W. J., Allen, A. K. & Cammue, B. P. A. A New Lectin from Meadow Saffron (*Colchicum autumnale*). *Plant Physiol.* **82**, 1036–1039 (1986).
 207. Kaku, H., Peumans, W. J. & Goldstein, I. J. Isolation and characterization of a second lectin (SNA-II) present in elderberry (*Sambucus nigra* L.) bark. *Arch. Biochem. Biophys.* **277**, 255–262 (1990).
 208. Kochibe, N. & Furukawa, K. Purification and Properties of a Novel Fucose-Specific Hemagglutinin of *Aleuria aurantia*. *Biochemistry* **19**, 2841–2846 (1980).
 209. Romano, P. R. *et al.* Development of recombinant *Aleuria aurantia* lectins with altered binding specificities to fucosylated glycans. *Biochem. Biophys. Res. Commun.* **414**, 84–89 (2011).
 210. Hirabayashi, J. & Arai, R. Lectin engineering: The possible and the actual. *Interface Focus* **9**, (2019).
 211. Chen, S., Qin, R. & Mahal, L. K. Sweet systems: technologies for glycomic analysis and their integration into systems biology. *Crit. Rev. Biochem. Mol. Biol.* **56**, 301–320 (2021).
 212. Haslam, S. M. *et al.* Chapter 50: Structural Analysis of Glycans. in *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2022). doi:10.1101/GLYCOBIOLOGY.4E.50.
 213. Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles - Part 1: Separation-based methods. *MAbs* **7**, 167–179 (2015).
 214. Hardy, M. R. & Townsend, R. R. Separation of positional isomers of oligosaccharides and glycopeptides by high-performance anion-exchange chromatography with pulsed amperometric detection. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3289–3293 (1988).
 215. Gennaro, L. A., Harvey, D. J. & Vouros, P. Reversed-phase ion-pairing liquid chromatography/ion trap mass spectrometry for the analysis of negatively charged, derivatized glycans. *Rapid Commun. Mass Spectrom.* **17**, 1528–1534 (2003).
 216. Zaia, J. Mass spectrometry and glycomics. *OMICS A Journal of Integrative Biology* vol. 14 401–418 (2010).
 217. Wuhrer, M. *et al.* Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics* **7**, 4070–4081 (2007).
 218. Rudd, P. M. *et al.* Chapter 51: Glycomics and Glycoproteomics. in *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2022). doi:10.1101/GLYCOBIOLOGY.4E.51.
 219. Powers, T. W. *et al.* Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked Glycan expression

- in tissues. *Anal. Chem.* **85**, 9799–9806 (2013).
220. Griffiths, J. A brief history of mass spectrometry. *Anal. Chem.* **80**, 5678–5683 (2008).
221. Dole, M. *et al.* Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240–2249 (1968).
222. Fenn, J. B. Electrospray ionization mass spectrometry: How it all began. *J. Biomol. Tech.* **13**, 101–118 (2002).
223. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **85**, 2–9 (2013).
224. Wilm, M. Principles of electrospray ionization. *Mol. Cell. Proteomics* **10**, (2011).
225. Lewis, J. K., Wei, J. & Siuzdak, G. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis. *Encycl. Anal. Chem.* 5880–5894 (2006) doi:10.1002/9780470027318.a1621.
226. Haag, A. M. Mass analyzers and mass spectrometers. in *Advances in Experimental Medicine and Biology* vol. 919 157–169 (Springer New York LLC, 2016).
227. Miller, P. E. & Denton, M. B. The quadrupole mass filter: Basic operating concepts. *J. Chem. Educ.* **63**, 617–622 (1986).
228. Mamyrin, B. A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int. J. Mass Spectrom.* **206**, 251–266 (2001).
229. Cotter, R. J. The New Time-of-Flight Mass Spectrometry. *Anal. Chem.* **71**, 445A–451A (2011).
230. Graham Cooks, R., Glish, G. L., McLuckey, S. A. & Kaiser, R. E. Ion trap mass spectrometry. *Chem. Eng. News* **69**, 26–41 (1991).
231. Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**, 1–35 (1998).
232. Koppelaar, D. W. *et al.* MS detectors. *Anal. Chem.* **77**, (2005).
233. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
234. Powers, T. W. *et al.* MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. *PLoS One* **9**, 1–11 (2014).
235. Blaschke, C. R. K. *et al.* Glycan Imaging Mass Spectrometry: Progress in Developing Clinical Diagnostic Assays for Tissues, Biofluids, and Cells. *Clinics in Laboratory Medicine* vol. 41 247–266 (2021).
236. Zavalin, A. *et al.* Direct imaging of single cells and tissue at sub-cellular spatial resolution using transmission geometry MALDI MS. *J. Mass Spectrom.* **47**, 1473–1481 (2012).
237. Good, C. J. *et al.* High Spatial Resolution MALDI Imaging Mass Spectrometry of Fresh-Frozen Bone. *Anal. Chem.* **94**, 3165–3172 (2022).
238. He, M. J. *et al.* Comparing DESI-MSI and MALDI-MSI Mediated Spatial Metabolomics and Their Applications in Cancer Studies. *Front. Oncol.* **12**, 1–13 (2022).
239. Kompauer, M., Heiles, S. & Spengler, B. Atmospheric pressure MALDI mass

- spectrometry imaging of tissues and cells at 1.4- μ m lateral resolution. *Nat. Methods* **14**, 90–96 (2016).
240. Fischer, J. L. *et al.* Matrix-Assisted Ionization-Ion Mobility Spectrometry-Mass Spectrometry: Selective Analysis of a Europium-PEG Complex in a Crude Mixture. *J. Am. Soc. Mass Spectrom.* **26**, 2086–2095 (2015).
 241. Bruderer, R. *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (2015).
 242. Mehta, A., Herrera, H. & Block, T. *Glycosylation and liver cancer. Advances in Cancer Research* vol. 126 (Elsevier Inc., 2015).
 243. Zhao, Y. *et al.* Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci* **99**, (2008).
 244. Wang, M. *et al.* Novel fucosylated biomarkers for the early detection of hepatocellular carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1914–1921 (2009).
 245. West, C. A. *et al.* N-Linked Glycan Branching and Fucosylation Are Increased Directly in Hcc Tissue As Determined through in Situ Glycan Imaging. *J. Proteome Res.* **17**, 3454–3462 (2018).
 246. Zhu, S. & Hoshida, Y. Molecular heterogeneity in hepatocellular carcinoma. *Hepatic Oncol.* **5**, HEP10 (2018).
 247. Maurice, J. & Manousou, P. Non-alcoholic fatty liver disease. *Clin. Med. J. R. Coll. Physicians London* **18**, 245–250 (2018).
 248. Zhao, T. *et al.* Heterogeneities of Site-Specific N-Glycosylation in HCC Tumors With Low and High AFP Concentrations. *Front. Oncol.* **10**, 496 (2020).
 249. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* vol. 291 2364–2369 (2001).
 250. Oda, Y., Hosokawa, N., Wada, I. & Nagata, K. EDEM as an acceptor of terminally misfolded glycoproteins released from calnexin. *Science (80-.).* **299**, 1394–1397 (2003).
 251. Hebert, D. N., Foellmer, B. & Helenius, A. Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell* **81**, 425–433 (1995).
 252. Wang, X. *et al.* Dysregulation of TGF- β 1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15791–15796 (2005).
 253. Paszek, M. J. *et al.* The cancer glycoalyx mechanically primes integrin-mediated growth and survival. *Nature* **511**, 319–325 (2014).
 254. Kailemia, M. J., Park, D. & Lebrilla, C. B. Glycans and glycoproteins as specific biomarkers for cancer. *Anal. Bioanal. Chem.* **409**, 395–410 (2017).
 255. Haglund, C. *et al.* Evaluation of CA 19-9 as a serum tumour marker in pancreatic cancer. *Br. J. Cancer* **53**, 197–202 (1986).
 256. Tan, P. S. *et al.* Clinicopathological indices to predict hepatocellular carcinoma molecular classification. *Liver Int.* **36**, 108–118 (2016).
 257. Hung, M. H. & Wang, X. W. Molecular Alterations and Heterogeneity in

- Hepatocellular Carcinoma. in *Hepatocellular Carcinoma: Translational Precision Medicine Approaches* (ed. Hoshida, Y.) 293–316 (Humana Press, 2019). doi:10.1007/978-3-030-21540-8_14.
258. Zhu, A. X. *et al.* Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α -fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **20**, 282–296 (2019).
 259. Singal, A. G., Pillai, A. & Tiro, J. Early Detection, Curative Treatment, and Survival Rates for Hepatocellular Carcinoma Surveillance in Patients with Cirrhosis: A Meta-analysis. *PLoS Med.* **11**, e1001624 (2014).
 260. Comunale, M. A. *et al.* Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *J. Proteome Res.* **5**, 308–315 (2006).
 261. Comunale, M. A. *et al.* Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J. Proteome Res.* **8**, 595–602 (2009).
 262. Pierce, M., Buckhaults, P., Chen, L. & Fregien, N. Regulation of N-acetylglucosaminyltransferase V and Asn-linked oligosaccharide $\beta(1,6)$ branching by a growth factor signaling pathway and effects on cell adhesion and metastatic potential. *Glycoconjugate Journal* vol. 14 623–630 (1997).
 263. Goossens, N., Sun, X. & Hoshida, Y. Molecular classification of hepatocellular carcinoma: potential therapeutic implications. *Hepatic Oncol.* **2**, 371–379 (2015).
 264. Lee, J.-S. S. *et al.* Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* **40**, 667–676 (2004).
 265. Kaposi-Novak, P. *et al.* Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. *J. Clin. Invest.* **116**, 1582–1595 (2006).
 266. Chaisaingmongkol, J. *et al.* Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell* **32**, 57-70.e3 (2017).
 267. Jiang, Y. *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261 (2019).
 268. Qiu, Z. *et al.* A Pharmacogenomic Landscape in Human Liver Cancers. *Cancer Cell* **36**, 179-193.e11 (2019).
 269. Hirschfield, H. *et al.* In vitro modeling of hepatocellular carcinoma molecular subtypes for anti-cancer drug assessment. *Exp. Mol. Med.* **50**, e419 (2018).
 270. Powers, T. W., Holst, S., Wuhler, M. & Mehta, A. S. Two-Dimensional N-Glycan Distribution Mapping of Hepatocellular Carcinoma Tissues by MALDI-Imaging Mass Spectrometry. *Biomolecules* **5**, 2554–2572 (2015).
 271. Ceroni, A. *et al.* GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659 (2008).
 272. Liu, X. E. *et al.* N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. *Hepatology* **46**, 1426–1435 (2007).
 273. Shultz, M. *et al.* The tumor-associated glycosyltransferase ST6Gal-I regulates stem

- cell transcription factors and confers a cancer stem cell phenotype. *Cancer Res.* **76**, 3978–3988 (2016).
274. Buck, C. A., Glick, M. C. & Warren, L. Glycopeptides from the surface of control and virus-transformed cells. *Science (80-.)*. **172**, 169–171 (1971).
 275. Wang, X. *et al.* Core fucosylation regulates epidermal growth factor receptor-mediated intracellular signaling. *J. Biol. Chem.* **281**, 2572–2577 (2006).
 276. Dennis, J. W., Laferté, S., Waghorne, C., Breitman, M. L. & Kerbel, R. S. β 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science (80-.)*. **236**, 582–585 (1987).
 277. Partridge, E. A. *et al.* Regulation of cytokine receptors by golgi N-glycan processing and endocytosis. *Science (80-.)*. **306**, 120–124 (2004).
 278. Drake, R. R. *et al.* MALDI Mass Spectrometry Imaging of N-Linked Glycans in Cancer Tissues. in *Advances in Cancer Research* vol. 134 85–116 (Academic Press Inc., 2017).
 279. Rich, N. E. *et al.* Hepatocellular Carcinoma Demonstrates Heterogeneous Growth Patterns in a Multicenter Cohort of Patients With Cirrhosis. *Hepatology* **72**, 1654–1665 (2020).
 280. Stowell, S. R., Ju, T. & Cummings, R. D. Protein glycosylation in cancer. *Annu. Rev. Pathol. Mech. Dis.* **10**, 473–510 (2015).
 281. Glavey, S. V. *et al.* The cancer glycome: Carbohydrates as mediators of metastasis. *Blood Rev.* **29**, 269–279 (2015).
 282. Furukawa, J. I. *et al.* Comprehensive glycomics of a multistep human brain tumor model reveals specific glycosylation patterns related to malignancy. *PLoS One* **10**, 1–25 (2015).
 283. Sweeney, J. G. *et al.* Loss of GCNT2/I-branched glycans enhances melanoma growth and survival. *Nat. Commun.* **9**, 1–18 (2018).
 284. Bones, J., Mittermayr, S., Donoghue, N. O. & Rudd, P. M. Ultra Performance Liquid Chromatographic Profiling of Serum N -Glycans for Fast and Efficient Glycosylation. *Anal. Chem.* **82**, 10208–10215 (2010).
 285. Lattová, E. *et al.* Mass spectrometric study of N-glycans from serum of woodchucks with liver cancer. *Rapid Commun. Mass Spectrom.* **23**, 2983–2995 (2009).
 286. McDowell, C. T., Lu, X., Mehta, A. S., Angel, P. M. & Drake, R. R. Applications and continued evolution of glycan imaging mass spectrometry. *Mass Spectrom. Rev.* 1–32 (2021) doi:10.1002/mas.21725.
 287. Blaschke, C. R. K., Black, A. P., Mehta, A. S., Angel, P. M. & Drake, R. R. Rapid N-Glycan Profiling of Serum and Plasma by a Novel Slide-Based Imaging Mass Spectrometry Workflow. *J. Am. Soc. Mass Spectrom.* **31**, 2511–2520 (2020).
 288. Black, A. P. *et al.* A novel mass spectrometry platform for multiplexed N-glycoprotein biomarker discovery from patient biofluids by antibody panel based N-glycan imaging. *Anal. Chem.* **91**, 8429–8435 (2019).
 289. Clift, C. L., Mehta, A., Drake, R. R. & Angel, P. M. Multiplexed Imaging Mass Spectrometry of Histological Staining, N-Glycan and Extracellular Matrix from One Tissue Section: A Tool for Fibrosis Research. *Methods Mol. Biol.* **2350**, 313–329

- (2021).
290. West, C. A., Liang, H., Drake, R. R. & Mehta, A. S. New Enzymatic Approach to Distinguish Fucosylation Isomers of N-Linked Glycans in Tissues Using MALDI Imaging Mass Spectrometry. *J. Proteome Res.* **19**, 2989–2996 (2020).
 291. Heimer, R. & Meyer, K. Studies on Sialic Acid of Submaxillary Mucoid. *Proc Natl Acad Sci U S A* **42**, 728–734 (1956).
 292. Varki, A. Diversity in the sialic acids. *Glycobiology* **2**, 25–40 (1992).
 293. Corfield, A. P., Veh, R. W., Wember, M., Michalski, J. C. & Schauer, R. The release of N-acetyl- and N-glycolloyl-neuraminic acid from soluble complex carbohydrates and erythrocytes by bacterial, viral and mammalian sialidases. *Biochem. J.* **197**, 293–299 (1981).
 294. Harvey, D. J. Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* vol. 5 1774–1786 (2005).
 295. O'Connor, P. B. & Costello, C. E. A high pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry ion source for thermal stabilization of labile biomolecules. *Rapid Commun. Mass Spectrom.* **15**, 1862–1868 (2001).
 296. Christensen, S. & Egebjerg, J. Cloning, expression and characterization of a sialidase gene from *Arthrobacter ureafaciens*. *Biotechnol. Appl. Biochem.* **41**, 225 (2005).
 297. Norton, P. *et al.* Development and application of a novel recombinant Aleuria aurantia lectin with enhanced core fucose binding for identification of glycoprotein biomarkers of hepatocellular carcinoma. *Proteomics* **16**, 3126–3136 (2016).
 298. Karamessinis, P. M. *et al.* Marked defects in the expression and glycosylation of α 2-HS glycoprotein/fetuin-A in plasma from neonates with intrauterine growth restriction. *Mol. Cell. Proteomics* **7**, 591–599 (2008).
 299. Rillahan, C. D., Brown, S. J., Register, A. C., Rosen, H. & Paulson, J. C. High-throughput screening for inhibitors of sialyl- and fucosyltransferases. *Angew. Chemie - Int. Ed.* **50**, 12534–12537 (2011).
 300. Schauer, R. & Kamerling, J. P. Chemistry, Biochemistry and Biology of Sialic Acids. in *New Comprehensive Biochemistry* vol. 29 243–402 (1997).
 301. Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. *Nature Reviews Immunology* vol. 7 255–266 (2007).
 302. DelaCourt, A. & Mehta, A. Beyond glyco-proteomics—Understanding the role of genetics in cancer biomarkers. in *Advances in Cancer Research* (Academic Press, 2022). doi:10.1016/bs.acr.2022.07.002.
 303. Delacourt, A. T., Liang, H., Drake, R. R., Angel, P. M. & Mehta, A. S. Novel Combined Enzymatic Approach to Analyze Nonsialylated N-Linked Glycans through MALDI Imaging Mass Spectrometry. *J. Proteome Res.* **21**, 1930–1938 (2022).
 304. Scott, D. A. *et al.* GlycoFibroTyper: A Novel Method for the Glycan Analysis of IgG and the Development of a Biomarker Signature of Liver Fibrosis. *Front. Immunol.*

13, (2022).

305. Carter, H. B. Prostate-specific antigen-based screening for prostate cancer evidence report and systematic review for the us preventive services task force. *JAMA - Journal of the American Medical Association* vol. 319 1866–1868 (2018).