Determination of Glycan Urinary Biomarkers in The Urine of COVID-19 Positive and Negative Individuals Using Bottom-up Glycomics

by

Agbor Eyonghebi Tanyi

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2023 by the
Graduate Supervisory Committee:

Chad R Borges, Chair
Jeremy Mills
Jia Guo

ARIZONA STATE UNIVERSITY

August 2023

ABSTRACT

Based on past studies, urinary glycan biomarkers have the potential to be used as diagnostic and prognostic markers for treatment purposes. This study brought into play the bottom-up glycan node analysis approach to analyze 39 urine samples from COVID-19 positive and negative individuals using gas chromatography-mass spectrometry (GC-MS) to determine potential urinary glycan biomarkers of COVID-19. Glycan node analysis involves chemically breaking down glycans in whole biospecimens in a way that conserves both monosaccharide identity and linkage information that facilitates the capture of unique glycan features as single analytical signals. Following data acquisition, the student t-test was done on all the nodes, but only four prominent nodes (t-Deoxyhexopyranose, 2,3-Gal, t-GlcNAc, and 3,6-GalNAc with respective p-values 0.03027, 0.03973, 0.0224, and 0.0004) were below the threshold p-value of 0.05 and showed some differences in the mean between both groups. To eliminate the probability of having false positive p-values, Bonferroni correction was done on the four nodes but only the 3,6-GalNAc node emerged as the only node that was below the newly adjusted p-value. Because sample analyses were done in batches, the Kruskal Wallis test was done to know if the batch effect was responsible for the observed lower relative concentration of 3,6-GalNAc in COVID-19 positive patients than in negative patients. A receiver operating characteristic curve (ROC) was plotted for the 3,6-GalNAc node and the area under the curve (AUC) was calculated to be 0.84, casting the 3,6-GalNAc node was a potential biomarker of COVID-19. 3,6-GalNAc largely arises from branched O-glycan

core structures, which are abundant in mucin glycoproteins that line the urogenital tract. Lowered relative concentrations of 3,6-GalNAc in the urine of COVID-19 positive patients may be explained by compromised kidney function that allows non-mucinous glycoproteins from the blood to contribute a greater proportion of the relative glycan node signals than in COVID-19 negative patients. Future prospective clinical studies will be needed to validate both the biomarker findings and this hypothesis.

DEDICATION

This Thesis is dedicated to my foster mother, the late Mrs. Ashu Regina Arrey Assolo, and to my uncle, the late Mr. Eyonghebi John Abangmanyi, and to my mother, Ms Agbor Ejeako, and to all my siblings for the individual roles and support they have played in my life that has led to this stage of my academic career.

ACKNOWLEDGMENTS

I would like to express my sincere and heartfelt gratitude to all those that supported me in one way or the other throughout my master's degree program. This thesis wouldn't have been possible without your distinguished contributions and invaluable effort.

Firstly, I would like to express my utmost gratitude to my supervisor, Dr. Chad R Borges for his exceptional guidance, humility, patience, expertise, and for making the environment conducive for learning. His insightful feedback, and ever ready to listen and accept the ideas of others, are the pivotal instruments that have shaped the outcome of my thesis. I am ever grateful for the opportunity he gave me in his lab so that I could share in his wealth of knowledge.

I would also like to thank my committee members, Dr Jeremy Mills and Dr Jia Guo whose feedback and suggestion during my technical review has shaped my ideas towards bringing a successful written thesis. I am grateful for taking time out of your busy schedule to be a part of my committee.

I am very thankful to the school of molecular sciences, for the immense support they offered me via the available resources that have been put in place to facilitate the completion of my thesis. Particularly, I am grateful to the Biodesign Institute, Borges Lab for providing me with a space to work and actualize my goals.

I would also like to thank my family, friends, classmates, and colleagues for their support and valuable discussions. The exchange of ideas and collaborative spirit has really enriched my learning experience and I am grateful for those moments.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

The outburst of the coronavirus disease (COVID-19) has taken a toll on the health of so many individuals. According to Statista, about 687 million people have been affected by this disease worldwide as of May 2nd, 2023[1]. For us to mitigate and curb the spread of this disease, it is imperative that we come up with accurate diagnostic and prognostic measures that will end the spread of such menace to our healthcare systems worldwide. Biomarkers have been identified as one of the potential tools that could be used for diagnoses and treatment purposes, including urinary glycan biomarkers[2]. Since urine is a readily available biological fluid, studying and analyzing the urinary biomatrix of Covid-19 positive and negative individuals will help us make significant verdicts for potential biomarkers that will be used for treatment purposes[3].

Urine Biomatrix: A Non-Invasive Source of Information

Urine is a biological waste product released from the system by the kidney through blood filtration. Urine is a complex biomatrix that consists of proteins, lipids, metabolites, and nucleic acids[4]. Improper filtration of the kidney which may be caused by kidney disease, urinary tract infection, diabetes, and inflammation, also contributes to its complexity[20]. The non-invasive nature of urine makes it stand out potentially as one of the most valuable and readily available sources of information about our health. Using urine for

diagnosis comes with many advantages because its non-invasive nature makes it easy for sample collection and very convenient for research purposes. One study stated that proteins in urine are highly glycosylated[15], which can either occur at asparagine residues (N-glycosylation) or on serine/threonine residues (O-glycosylation). Glycosylation is a post-translational modification process that adds complex, polymeric sugar moieties to substrates such as proteins and lipids. It plays a significant role in protein folding and signaling. A recent study by the Yale Impact team[16] found out that SARS-CoV-2 spike protein is present in the urine of Covid-19 patients. The causative agent of the coronavirus disease, SARS-CoV-2 contains spike proteins that bind to cell surfaces through the angiotensin-converting enzyme 2[17] (ACE2). This ACE2 is a glycoprotein that is involved in glycosylation regulation processes[17].

Many biofluids such as blood, plasma, and extracellular vesicles have been used to study and carve out glycosylation-based biomarkers for future diagnostic and prognostic purposes. Plasma-derived extracellular vesicles were used by the Borges group to ascertain cancer biomarkers[12]. Urine too has been used to determine urinary biomarkers, and this was done by Xu, M[15] et al. In their research, they used urine to determine glycoprotein biomarkers.

Fundamental Biochemistry of Monosaccharides, Carbohydrates, And Glycans.

Carbohydrates are macromolecules that consist of carbon, hydrogen, and oxygen. They are commonly found in nature. They have a general formula, $C_y(H_2O)n$, where "y" represents the number of carbon atoms, and n is an integer in the range of 3-9. They possess a carbonyl group, an aldehyde, or a ketone. Due to their sweet nature, they are referred to as saccharides or sugars and act as important sources of energy for all living organisms. Carbohydrates are generally categorized into three main groups: monosaccharides, disaccharides, and polysaccharides. Monosaccharides are the simplest of the carbohydrate family of compounds, they constitute the fundamental units from which all other classes of carbohydrates are built. When several monosaccharide monomers are covalently bonded together, they give rise to polysaccharides or polyhydroxy alcohols. They are sometimes called polyhydroxy alcohols because they contain multiple hydroxyl groups in their structure.

Emil Fischer, a German chemist, first spearheaded the classification of all monosaccharides in what is today known as the Fischer projections. In his classification, he noted that all monosaccharides have a long chiral backbone of hydroxymethylene units that begins with either an aldehyde or a ketone and terminates with a hydroxymethyl group (CH_2OH). Glycan monomers in solution, exist as an equilibrium mixture of both cyclic and acyclic forms due to a phenomenon known as mutarotation. In

solution, they prefer the cyclized form because they are thermodynamically stable in that form and exist in lower concentrations in the open chains. The open chain form is an aldehyde while the cyclized form is a hemiacetal. The hemiacetal can either be a five (furanose) or six (pyranose) member ring. A six-member ring or pyranose is formed when the hydroxyl group (OH) at carbon five reacts with an aldehyde at the anomeric carbon (C-1), and its stereo configuration is determined depending on the position of the OH group. When the OH group at carbon one in the newly formed ring is above the plane of the ring, it is denoted a beta pyranose but when its below is denoted an alpha pyranose. For the furanoses, the carbon four hydroxy group reacts with an aldehyde at the anomeric carbon (C-1) to produce a furanose ring, and its stereo-configuration is determined by the orientation C-1 OH group on the newly formed ring as explained above.

Glycans are a class of biopolymers that are made up of several monosaccharide units linked by glycosidic bonds. Glycosidic bonds are covalent bonds that link the hemiacetal alcohol of one sugar molecule to the hydroxyl group of another. Figures 1 and 2 below provide an idea of what glycosidic bonds and various glycan types look like. Glycans are present in all living organisms where they offer structural support, energy storage and metabolism, cell-cell signaling, and transport function. They are also referred to as polysaccharides or oligosaccharides because they are carbohydrate-based polymers. Just like polysaccharides, they are constructed by either linear or branched chains of

monosaccharide residues linked by glycosidic bonds. Glycan assembly takes place within cells, and they are synthesized in the endoplasmic reticulum (mostly N-glycans) and Golgi apparatus (mostly O-glycans). Glycans are complex polysaccharides because the monosaccharide moieties involved in their construction might have different monomeric building units that exist in different isomeric forms and can be linked together in different ways—including in both linear and branched formats. When it comes to protein-linked glycans, there are two major classes: N-linked glycans and O-linked glycans, described in greater detail below.

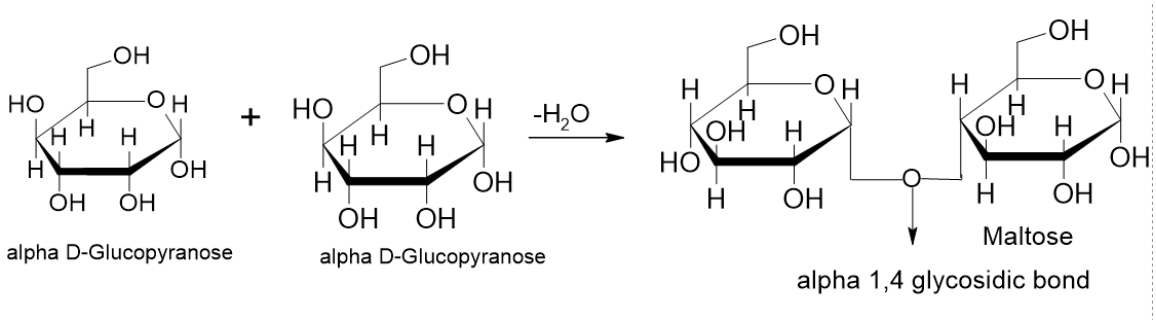

Figure 1: The formation of maltose from two glucose monomer units, illustrating an alpha-1,4-glycosidic bond.

Biosynthetic Structures of Glycans.

Glycan biosynthesis takes place in the endoplasmic reticulum and Golgi apparatus. The biosynthesis of glycans is generally determined by the enzyme known as glycosyltransferase. These enzymes are responsible for the building up/construction of monosaccharides into linear and branched chain polymers, although they do so in non-template driven manner unlike proteins, DNA, and RNA that have specific instructions and directions on how they are built. Glycosyltransferase is an embodiment of enzymes that catalyze reactions involving the transfer of a monosaccharide residue of a nucleotide sugar donor (UDP-Gal, GDP-Fuc, or CMP-Sia) to an acceptor substrate such as proteins or lipids. But prior to their synthesis, the sugar nucleotides are activated by an enzyme known as hexokinase. Hexokinase adds a phosphate group to the first carbon of the sugar molecule thereby making it ready for the construction of glycan chains. Glycosyltransferases grow glycan chains in a successive manner such that the product/outcome of one enzyme makes it a perfect substrate for another.

Glycosyltransferase enzymes (GT) exhibit a high degree of specificity, meaning the enzyme can only catalyze reactions that add sugar moieties to a recognized acceptor substrate with specific linkage instructions. Some major GTs and the reactions they catalyze include; **galactosyltransferase (GalT)**[5], which adds galactose monosaccharide

residues to different acceptor molecules that include glycoproteins and glycolipids, **Fucosyltransferase** adds fucose residues to glycoproteins and glycolipids including N and O glycans, **Sialyltransferase** catalyze biological reactions that add sialic acid (Neu5A) residues to glycoproteins and glycolipids to produce sialylated glycans, **N-acetylglucosaminyltransferase (GMT)** adds N-acetylglucosamine (GlcNAc) residues to the interior structure of N-glycans, creating more complex structure, **Glucoronyltransferases** which adds glucuronic acid (GlcA) residues to different acceptor molecules, including glycosaminoglycans (GAGs) and proteoglycans**.** Irregular glycosyltransferase expression can produce complex heterogenous aberrant glycans[11], known as altered glycosylation.

Glycans are classified into different types. Glycans attached to proteins are generally classified as **N-Glycans, O-Glycan, and Proteoglycans. N-glycans** are formed by the covalent attachment of a sugar molecule to the nitrogen of an asparagine side chain in a protein molecule through a process called N-glycosylation, and the bond that exists between them is an N-glycosidic bond. N-glycans are synthesized in the endoplasmic reticulum and Golgi apparatus. They are located on a variety of proteins such as cytoplasmic proteins, membrane proteins, secreted proteins, and have a consensus sequence as Asn-X-Ser/Thr where X can be any amino acid except proline. **O-linked glycans** are formed by covalently bonding their sugar molecules to the oxygen atom of a

serine or threonine residue in a protein molecule. O-glycans are synthesized in the Golgi apparatus and have no consensus sequence. **Proteoglycans** are glycans bonded to proteins via the glycosaminoglycan (GAG) chain. GAGs are found in the extracellular matrix and cell membrane where they provide structure and support to tissues and organs. They are not really considered as glycans because they are long straight-chain monosaccharides units linked to proteins.

The SARS-CoV-2 in coronavirus has spike proteins that are glycosylated[6], and it uses the spike protein as a means of attaching itself to the body cells. When this happens, inflammatory proteins are produced which leads to cell damage. Using glycans as biomarkers for diseases including Covid-19 may be promising because they exhibit a high degree of specificity during their construction and as such can be very specific to different viruses. With this trait, it will be easier to know what is responsible for a particular infection, then turn it into a possible biomarker.

N-linked glycosylation **O-linked Glycosylation (Core-2 type)**

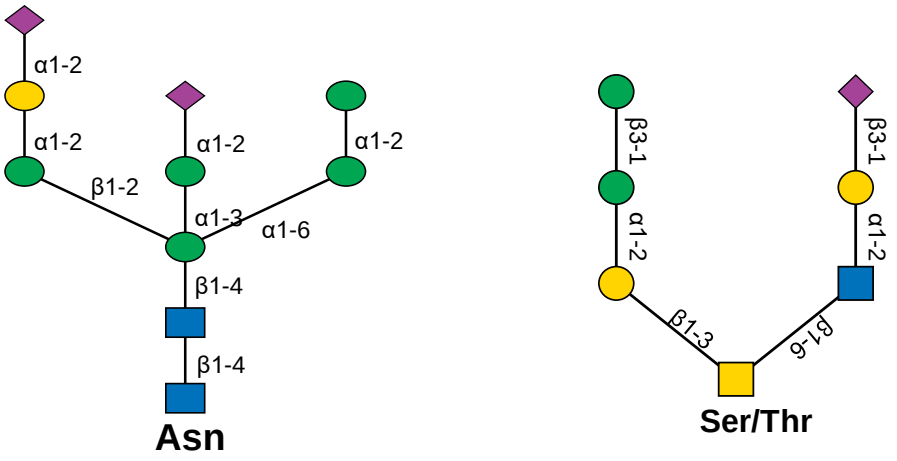

Figure 2: Major glycan types (N-linked and O-linked)

The meaning of the monomers colored shaped can be found in the legend of figure 5.

Gas Chromatography (GC):

Chromatography is an essential analytical separations technique used in glycobiology to study glycans. Since glycans are complex biopolymers, they require specialized chromatographic techniques for their separation and analysis.

The GC is a common chromatographic technique used in analytical research for separating, analyzing, and quantifying volatile organic compounds in a complex mixture. The GC operates by injecting volatile liquid sample into a mobile phase which is a carrier gas such as nitrogen or helium. The mobile phase then transports the sample to a column packed with a liquid or solid stationary phase material. The sample components are then separated depending on their physicochemical properties and the degree of interaction with the stationary phase, a thin film of liquid in most cases. GC can only be used to analyze glycans after they have been broken down into their constituent monosaccharides and derivatized to enhance their volatility and diminish their polarity.

Apart from GC, some of the chromatographic techniques used for analyzing glycans are;

a. High-performance liquid chromatography (HPLC)[7].

   HPLC can be used to separate and analyze complex glycans based on their charge, size, and other properties. The stationary phase can be a stationary medium, while the mobile phase is a liquid.

b. The Ion Exchange Chromatography (IEC) [8]

The IEC is another chromatography technique used in glycobiology to separate glycans based on their net surface charge. The stationary phase here is equipped with charged groups that interact with the ionic functional group of glycans in the mobile phase.

c. Gel-filtration Chromatography[9]

This chromatography technique is used to separate biomolecules based on their different molecular sizes. Here, compounds can be separated by their sizes via a porous gel stationary phase and a liquid mobile phase.

d. Lectin Affinity Chromatography[10]

It is instrumental in glycan analysis. It works on the principle of isolating and purifying specific glycans types. The stationary phase contains lectins, proteins by nature that bind to specific types of glycans. The glycan of interest is then bound to the lectin and eluted for further analysis.

Mass Spectrometry (MS): An essential tool in glycobiology

The importance of the mass spectrometer today, especially in analytical chemistry, cannot be overemphasized. It is a powerful tool that measures the mass-to-charge ratio of ionizable biomolecules in a sample. The mass spectrometer can be used to determine the size and structure of unknown

compounds and quantify known ones. There are three main components that act as the backbone behind the working principle of the mass spectrometer. All have an ionization source, mass analyzer, and detector. For samples to be analyzed in the mass spectrometer, they must be in the gaseous phase and ionized to facilitate manipulation by the electric and/or magnetic field(s). The ions pass through the mass analyzer where they are sorted out and separated according to their mass-to-charge ratio (m/z). The separated ions go through the detector that records them as current and outputs them as digital signals or a mass spectrum. The mass spectrometer is an essential tool used in glycobiology to separate and quantify biomolecules of interest such as glycans, proteins, nucleic acids. With mass spectrometry, the intact mass and fragments, linkage, and branching pattern can be determined. Different mass spectrometric ionization techniques can be used to analyze and quantify intact glycans. Some of the common ones include Matrix-assisted laser desorption ionization (MALDI), and electrospray ionization (ESI). MALDI is a soft ionization technique that uses a laser absorbing-type matrix to create gas phase ions from large biomolecules such as glycans to analyze and quantify them. While ESI uses mass spectrometry to analyze large ionizable biopolymers. Since large, intact glycans were not the analytical end-product in this study, neither MALDI nor ESI were used. Rather, electron ionization (EI) was

used to ionize the small, gas-phase glycan monosaccharide derivatives that eluted from the GC. EI provides an analytical advantage of producing highly reproducible fingerprint-like mass spectra from small molecules that can be quite helpful for identification purposes.

Glycan Node Analysis (GNA): A molecularly Bottom-up Approach

"Glycomics" means studying the complete collection of glycans (glycome) in a biospecimen, whether gratis or present in complex mixtures, produced by a cell or tissue under specific conditions such as time, location, and environment. Bottom-up glycomics is used to break down macromolecules such as glycoproteins using enzymatic or chemical methods into smaller pieces (oligosaccharides) to obtain information about their linkage pattern. These smaller pieces can then be analyzed using mass spectrometry and other proven analytical methods. The sole objective of this approach is to obtain detailed information on glycan structure, monosaccharide compositions, linkage types, sequence, and branching patterns. With this approach, researchers have developed diagnostic and therapeutic applications that explain the role of glycans in many biological processes. Although the bottom-up glycomics approach has proven advantageous, several challenges are associated with it, and glycan node analysis (GNA) is one of the sub-techniques of this approach that will be used to demonstrate these advantages and its challenges. One of the challenges faced by GNA as the first molecularly bottom-up

approach to glycomics is that one can lose information on intact glycan structure, thereby leaving us clueless on what the mass of the original polymer was. Also, some glycans may resist enzymatic or chemical actions, which can truncate the scope of our analysis and data interpretation. For example, some monosaccharide functional groups such as carboxyl, amines, sulfates, and phosphates have charged functional groups at physiological pH that directly impede their detection with the bottom-up glycomics approach. In 2013, the Borges group demonstrated the efficiency and reliability of glycan node analysis as a bottom-up approach by conducting glycan methylation analysis (a.k.a. linkage analysis) on different biofluids such as blood, plasma, serum, seminal fluid, saliva, urine[13]. Their main objective was to use glycan node analysis (GNA), a bottom-up approach, as a tool for biomarker verdicts for clinical diagnosis and prognostic purposes to detach glycans in their complex biomatrices and obtain information about monosaccharides linkage pattern (glycan nodes).

Information about the various classes of glycans and their unique features can be captured simultaneously as single analytical signals through the following glycan node analysis procedures that were adapted from Heiss et al[14] (Fig. 3 and 4): Firstly, the biospecimen will be permethylated to introduce a methyl substituent to replace any available hydrogen attached to oxygen or nitrogen on the original glycan polymer. At this point, the O-linked glycans are released, followed by liquid/liquid extraction of the

permethylated glycans and glycoproteins to separate them from other sample component mixtures. The permethylated glycans are then hydrolyzed with trifluoroacetic acid (TFA) to break glycosidic linkages between major glycans into their constituent monosaccharides and introduce hydroxyl groups at positions where there was formerly a linkage. The hydrolyzed glycans are then reduced with sodium borohydride thereby converting the aldehydes into alditols (alcohols). Reducing them helps to prevent mutarotation which would complicate subsequent chromatographic separations and the resulting chromatograms. Finally, all the free hydroxyl groups are then acetylated, producing partially methylated alditol acetates, which makes the monosaccharides volatile and easy to analyze by the GC-MS while conserving both monosaccharide composition and linkage information. A conceptual illustration of how this approach can be used to identify cancer biomarkers can be seen in Figure 5 below. In the same 2013 paper by the Borges group, they qualitatively profiled glycans from urine using the bottom-up glycomics approach. However, they did not go into much detail as far as determining what specific nodes are in urine that might serve as possible biomarkers for future diagnosis and treatment purposes.

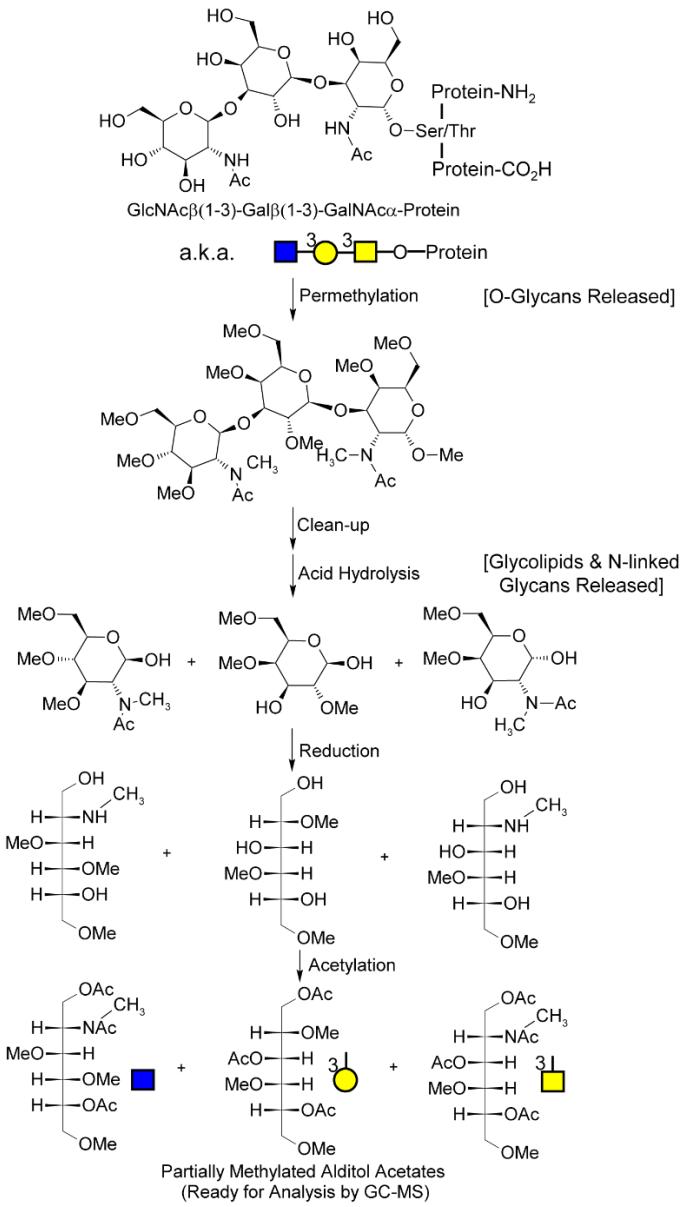

Figure 3: Molecular overview of the chemistry of glycan node analysis process[11]

This figure was reprinted with Permission from the American Chemical Society

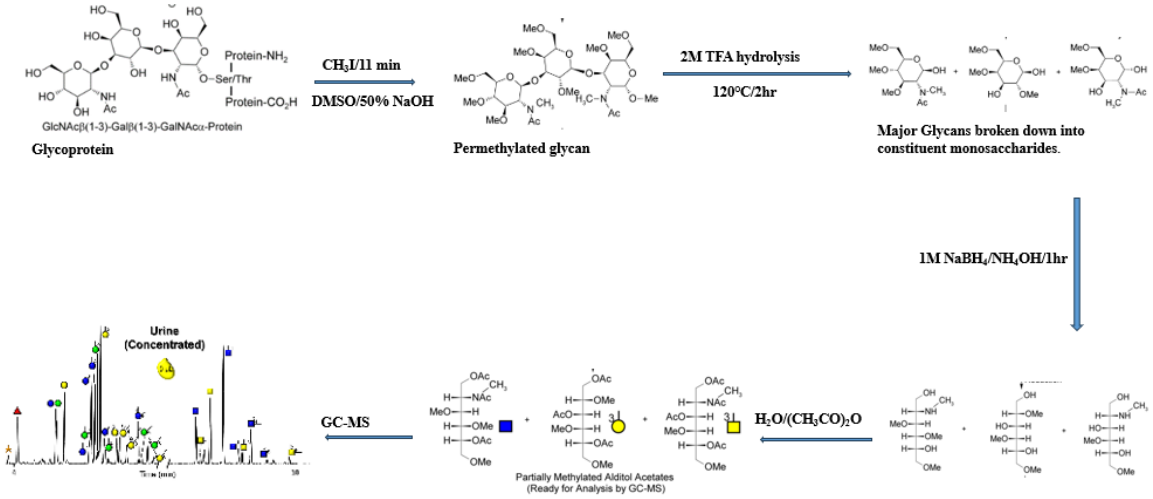

Figure 4: GNA Reaction Scheme

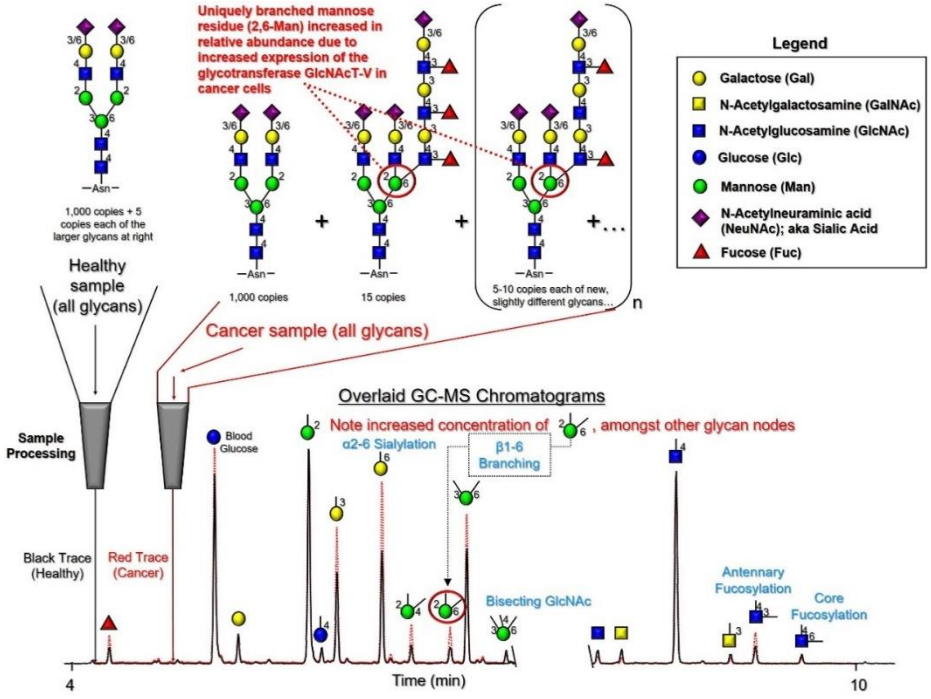

Figure 5. Conceptual overview of the bottom-up glycan node analysis approach that shows an overlaid chromatogram of a healthy individual and a cancer patient. This diagram shows an abberant glycan profile of a cancer patient that grows another arm of glycan due to the over expression of the glycosyltransferase enzyme, GlcNAcT-v in cancer cells. The chromatographic peaks in black depicts the healthy individual while the increased peak in red depicts that of a cancer patient. This image was reprinted with permission from the American Chemical Society.

Thesis Objective: Urinary glycans as biomarkers.

Since glycans play a key function in many biological processes and changes in their structure may be attributable to disease. The main objective of this study was to identify specific glycan features in the urine of Covid-19 positive individuals that can serve as possible biomarkers for identifying COVID-19 positive patients. To achieve this, the bottom-up glycan node analysis approach adopted from Dr. Borges' lab was used. The approach involves sample preparation, the standard GNA procedure, analyzing the samples using the GC-MS, data processing and data analysis.

CHAPTER 2

MATERIALS AND METHODS

Materials and Reagents

About 39 urine samples of Covid-19 Positive and Negative individuals (Table 1) were made available from the Mayo Clinic, Florida biobank. All specimens were collected for purposes other than this study. Specimens were collected by Mayo clinic personnel with informed consent, IRB approval, and in compliance with the declaration of Helsinki principles. Specimens were coded and de-identified prior to shipment ASU. ASU researchers will not have and will never gain access to the key/code link samples/data back to individual donors. Moreover, ASU personnel will never attempt to re-identify samples/data in the future. As determined by an ASU Office of Research Integrity and Assurance compliance officer, these conditions place this study in the category of non-human subjects research. Altogether, there were 20 positive and 19 negative covid-19 samples from individuals in the range of 36-99 years (Table 1 below).

Acetone (ACS reagent) 99.5% from Baker with Cat number 9006-33, USA; Sodium hydroxide beads, 97% reagent grade from Sigma-Aldrich, USA; Acetic Anhydride 99.5% reagent grade Sigma Aldrich with Cat number, 39996-1KG, U.S.A.; Dimethyl Sulfoxide Sigma-Aldrich with Cat number D8418-1L, USA; Methanol Fisher Chemicals with Cat number A456-4 from, USA; Ammonium hydroxide ACS reagent grade, 28-30% ammonia content, Sigma-Aldrich ,Cat number 221228-500ml-A, USA; Sodium

Borohydride, from Fluka Analytical, Sigma-Aldrich, Cat number 71321-100G, USA; Sodium phosphate monobasic (> 99.0%) Sigma-Aldrich, Cat number S8282-500G, USA; Sodium phosphate dibasic (> 99.0%) Sigma-Aldrich, with Cat number 37907-500G,, USA; Sodium chloride, from Sigma-Aldrich with Cat number S7653-1KG, USA; Iodomethane from Sigma-Aldrich Co Spruce street, st louis; SpeedVac concentrator from Savant SC250EXP; Compressed Nitrogen from Matheson, cat number ON1066 for drying; 1.5ml Eppendorf tube; Glycan node dedicated syringe; salinized glass test tube; Pocket refractometer from Atago ,with Cat number 4410, Japan; Trifluoroacetic acid (TFA) 99% reagent grade, from Sigma-Aldrich Cat number 299537-100G, USA; Methylene chloride from Fisher chemicals with Cat number D143-4, USA; Chloroform from Sigma Aldrich ,with Cat number 650498-4L, U.S.A.

Specific Gravity (SG): The instrument used in measuring the specific gravity was the pocket refractometer, cat number 4410, manufactured by ATAGO, Japan. About 110ul of each sample (Table 1) was placed in a 1.5ml Eppendorf tube. Then 10ul was aliquoted from it and placed on the screen of the calibrated refractometer, and its specific gravity value was recorded. SG test was done for each sample, and their specific gravity values were recorded before and after pre-concentration**.**

Pre-concentration

After their SG values were obtained, the samples were pre-concentrated using a speed Vac concentrator for about 35-40 min at 45℃ and three torr of pressure. Urine samples

were concentrated 3-fold, i.e., from 100 up to about 33 ul at each run. Afterward, samples were placed back at -80℃ or processed immediately.

Glycan Node Analysis (GNA)

The following glycan node procedure was adapted from Borges et al.[11]–whereby urine samples from Covid-19 positive and negative individuals were thawed, concentrated 3-fold, and then subjected to liquid phase permethylation as described below.

Permethylation/Purification of Glycans

The first step involved preparation of 2.5-2.6 ml of permethylation base in which 100ul from a freshly prepared 50% (w/v) NaOH solution was mixed with 200ul methanol to give a methanolic gelatinous methanolic base. The gelatinous methanolic base is then extracted with 4ml dimethyl sulfoxide (DMSO) and centrifuged at 3500 rcf for 1 min. This was done six times to consecutively remove any sodium (bi)carbonate found on top of DMSO and the excess DMSO. The final permethylated base was then reconstituted with 2 ml of DMSO (added slowly without bubbles) to have a total volume of 2.5-2.6 ml. After the permethylation base was fully prepared, 10ul of each urine sample was placed in a 1.5ml Eppendorf tube, reconstituted with 200ul of DMSO, and vortexed to ensure

proper mixing. Two hundred and thirty microliters of the permethylated base were added to the mixture (urine + DMSO), quickly followed by vortexing for 10s. The urine and DMSO mixture were incubated at room temperature for 4 min. After 4min, 100ul of Iodomethane was added with the help of a dedicated syringe and incubated again for 11 min, with intermittent stirring. The sample solution was transferred from a 1.5ml Eppendorf vial to a silanized 13 x 100 mm glass test tube containing 3.5ml 0.2M sodium phosphate buffer with 0.5M NaCl solution at pH 7; this was thoroughly mixed. 1.2ml of chloroform was added to the buffered urine samples, and proper mixing was ensured by shaking the glass tube, ensuring it was well capped. Then centrifugation, at 3500 rcf for 1 min, separated the top aqueous layer from the bottom organic layer. The top aqueous layer was discarded into waste, and the phosphate buffer was again replaced. Liquid-liquid extraction was done repeatedly three times to enhance the recovery of glycans from the complex glycan protein mixture. After three liquid/liquid (L/L) extraction rounds, the bottom organic layer (chloroform) containing the analyte was recovered by placing it on a metal block with a temperature set at 74℃ and dried under nitrogen gas.

Trifluoroacetic Acid (TFA) Hydrolysis and Reduction

After L/L extraction was done, the dried extracts were hydrolyzed by adding 325ul of 2M TFA to each sample. The urine samples were securely capped to prevent evaporative losses and fully wrapped with an aluminum foil while incubating for 2 hours at 120℃. After 2 hours, samples were centrifuged to bring down the liquid from the wall of the

tube, then allowed to dry under a gentle stream of nitrogen gas while seating on a heated metal block set at 74℃. The residue from TFA were chemically reduced by the addition of 475ul of freshly prepared sodium borohydride in 1M ammonium hydroxide to each sample; vortexing and sonication were done to ease the dissolution of the sample, followed by 1 hour incubation period at room temperature. To reduce the amount of borate, 63ul methanol was added, which was well mixed, vortexed, centrifuged, and dried under a gentle stream of nitrogen gas while set in a metal block set at 74℃. Then 125ul of methanol/acetic acid in a molar ratio of 9:1 was added, centrifuged, and dried similarly. Samples were thoroughly dried in a vacuum desiccator for about 15 min and could be stored at -80℃, depending on whether one wants to continue**.**

Acetylation (final step)

To acetylate hydroxyl groups from the reduced, partially methylated monosaccharides, about 18ul of deionized water was added to each glass tube to dissolve any precipitate. Then, 250ul of acetic anhydride was added into the tube; this was capped and vortexed thoroughly for 20s, followed by 2min sonication. After sonication, the samples were incubated for 10 min at 60℃ and then centrifuged to cool and bring down liquid particles from the tube side walls. To each glass tube was added 230ul conc TFA; this was centrifuged and incubated for 10 min at 50℃. Cleaning up the sample mixture involves liquid/liquid extraction, and this was done with 1.8 ml dichloromethane and two rounds deionized water (2 ml) washes for each sample. The top aqueous layer was discarded

each round into waste. The bottom organic layer was then transferred into a silanized auto-sampler vial with the help of a silanized pipette, dried under a gentle stream of nitrogen, and reconstituted in acetone as described below for injection onto the GC-MS.

GC-MS

A Waters 7890A chromatograph fitted with a CTC PAL autosampler injector was connected to a Waters Time-Of-Flight mass spectrometer used for this experiment. Before any analysis is done, the instrument was calibrated with perfluorotributylamine (PFTBA) to ensure accurate and reproducible results. After calibration, acetylated residues in individual vials to be analyzed are reconstituted with 30-40ul of fresh acetone, vortexed, then placed on an autosampler rack. Approximately 1ul of the 30-40ul of the total volume was injected at a split ratio of 20:1 (standard) or 15:1, depending on the analyte's concentration level in the solution sample. This injection was done on a hot silanized glass liner embedded with a small plug of silanized glass wool. Using helium as the carrier gas with a volumetric flow rate of 0.8ml/min, the volatilized samples were separated into a 30 m DB-ms column. 165℃ was the starting temperature of the GC, then there was a subsequent increase in temperature to 265℃ and started ramping to 325℃ at a rate of 30℃/min. The ramped temperature was finally kept stable at 325℃ for 3 min. As the separated GC components entered the TOF mass spectrometer, they were subjected to a 70eV electron ionization. For every 0.1s, the positive ion mode pulses generated by the TOF within a mass range of 400-800 m/z were summed.

Data Processing

Acquired data were transported to another computer with Masslynx installed on it with the help of a USB flash drive. Masslynx is a data processing software that is used to carry out data processing on all these glycan nodes. Then, the sum extracted ion chromatographic (XIC) peak areas of all glycan nodes were automatically integrated with the help of Quanlynx, which is integrated into the MassLynx software (To have an idea of what XIC corresponds to a particular node kindly check supporting information Table S1 in the 2013 paper published by Borges et al on multiplexed surrogate analysis of glycosyltransferase activity in whole blood plasma[13]). The summed extracted ion chromatograms (examples shown in figures 7 and 8) were automatically integrated (with manual verification) so that the total area occupied by the peaks could be known. The area is proportional to the concentration of the partially methylated alditol acetate (PMAA or simply "node") analyte found in the sample. The QuanLynx software sometimes omits to integrate some sample peaks. When this is identified, the peaks of the sample will be manually integrated to ensure all areas of such omitted peaks are captured. After all automatic and manual integration was done, the extracted ion chromatogram was referenced to the complex carbohydrate research center database (CCRC)[19] spectral database. To ensure quantitative precision of the processed data, the hexoses were normalized to all sum of endogenous hexoses, while the HexNAcs were normalized to all

sum of endogenous HexNAcs. Once they have been normalized, relative signals for all hexoses and HexNAcs were tabulated.

Data analysis

Glycan node profiles of Covid-19 positive patients were initially compared to those of Covid-19 negative patients with the help of the student's t-test in Microsoft Excel. To account for multiple comparisons, $p < 0.05/n$ glycan nodes where $n = 20$ was the p-value threshold applied for initial statistical significance assessment. The student t-test was done on all nodes for the two-group data (Covid-19 positive and negative) to know if there was any significant difference between the mean of the two data groups. Four p-values (Table 2) of some nodes (both Hexoses and HexNAcs) were below the set threshold p-value, i.e., 0.05. To reduce the likelihood of having false positive p-values, the Bonferroni correction was performed on the four p-values to have an adjusted p-value that will be used to compare the four p-values. The Bonferroni correction was done by dividing the threshold p-value to the total number of nodes (only nodes above 0.1% relative abundance). The adjusted p-value was calculated to be 0.0025. Also, the Kruskal Wallis test was done to check if the batch effect was responsible for any significant differences observed between positive and negative COVID-19 patients.

CHAPTER 3

RESULTS AND DISCUSSIONS

Results

Individual's Information and Sample Processing Logistics

Altogether there were 20 positive and 19 negative covid-19 samples from individuals in the range of 36-99 years (Table 1) below**.**

***Table 1: Samples utilized for this research and their processing order*.**

S/N	Covid-19 positive Identifiers	Covid-19 positive individual's age	Covid-19 positive gender	Covid-19 negative identifiers	Covid-19 negative individual age	Covid-19 negative gender	1st processed set	2nd processed set after randomization	3rd processed set after randomization	4th processed after randomization
1	COVP0608	99	Female	COVN0371	98	F	COVN0350	COVN0292	COVP0577	COVN0481
2	COVP0588	98	Male	COVN0132	95	F	COVN0681	COVN0293	COVP0608	COVP0593
3	COVP0572	94	Female	COVN0076	92	F	COVN0682	COVN0313	COVP0770	COVN0076
4	COVP0145	93	Female	COVN0313	91	M	COVNO322	COVN0371	COVN0083	COVP0607
5	COVP0709	90	Male	COVN0452	90	M	COVN0331	COVN0420	COVN0432	COVP0145
6	COVP0712	93	Female	COVN0293	89	F	COVN0638	COVP0712	COVN0004	COVP0559
7	COVP0726	90	Male	COVN0420	86	M		COVP0751	COVP0674	COVP0169
8	COVPO607	85	Male	COVN0004	85	M		COVPO572	COVPO709	COVPO593
9	COVP0751	65	Female	COVN0292	69	M		COVP0591	COVN0208	COVP0765
10	COVP0674	64	Female	COVN0481	69	M			COVNP0565	
11	COVP0577	68	Male	COVN0565	69	F			COVN0132	
12	COVP0593	68	Male	COVN0083	68	F			COVP0768	
13	COVPO599	67	Female	COVN0322	59	F			COVP0726	
14	COVP0468	59	Male	COVN0331	59	M			COVP0553	
15	COVP0768	60	Female	COVN0350	59	M			COVP0468	
16	COVP0591	56	Female	COVN0681	58	F			COVP0452	
17	COVP0770	56	Male	COVN0638	48	M			COVP0719	
18	COVP0765	43	Male	COVN0682	48	M				
19	COVP0169	40	Male	COVN0208	46	F				
20	COVP0553	37	Female	COVN0432	44	F				
21	COVP0719	36	Female							

Processed glycan node data were evaluated for significant differences in each node between COVID-19 positive and negative patients using a student's t-test (Table 2).

Table 2: Four glycan nodes that were lower than the set threshold student t-test p-value of 0.05. The one p-value that was lower than the Bonferroni-corrected threshold of 0.0025 is highlighted in green.

Glycan Nodes	Student t-test P-values
t-Deoxyhexopyranose	**0.03027**
2,3 Gal	**0.03973**
t-GlcNAc	**0.0224**
3,6 GalNAc	**0.0004**

Comparing the adjusted p-value with the student p-value for the four chosen nodes, it was realized that only the 3,6 GalNAc ($p = 0.0004$) node was below the Bonferroni-adjusted p-value, and the node became a focused point of study for this project.

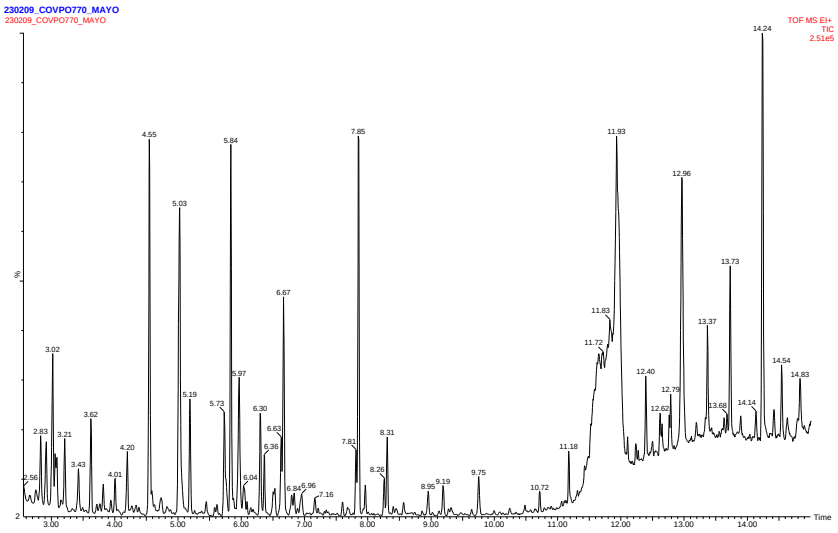

Figure 6: Illustrative total ion current chromatogram (TIC) from a urine glycan node analysis run.

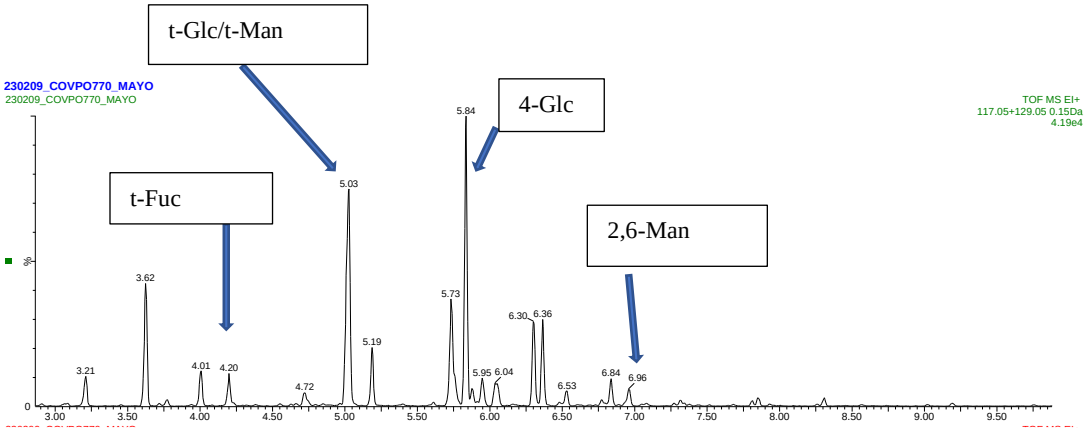

Figure 7: An illustrative XIC for hexoses (m/z 117.05 + 129.05), with a few of them annotated.

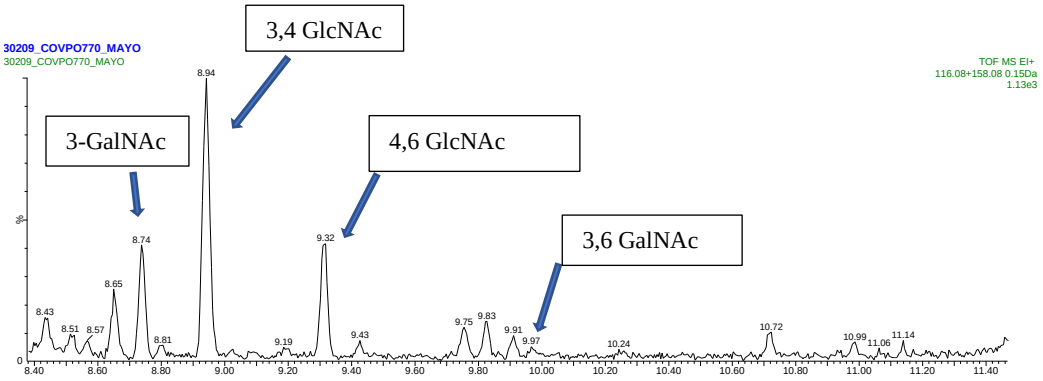

Figure 8: An illustrative XIC for HexNAcs (m/z 116.08 + 158.08) with a few of them annotated.

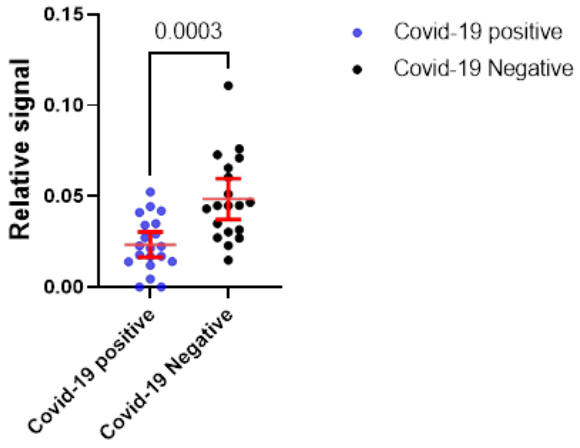

Figure 9: A univariate column scatter plot for 3,6 GalNAc

Figure 9 shows a univariant column scatter plot of the 3,6 GalNAc node for Covid-19 Positive and Covid-19 negative individuals with a 95% confidence interval (CI) and error bars.

A univariant scatter plot was created for 3,6-GalNAc to better understand the relative abundance of this node and its distribution in all the COVID-19 positive and negative.

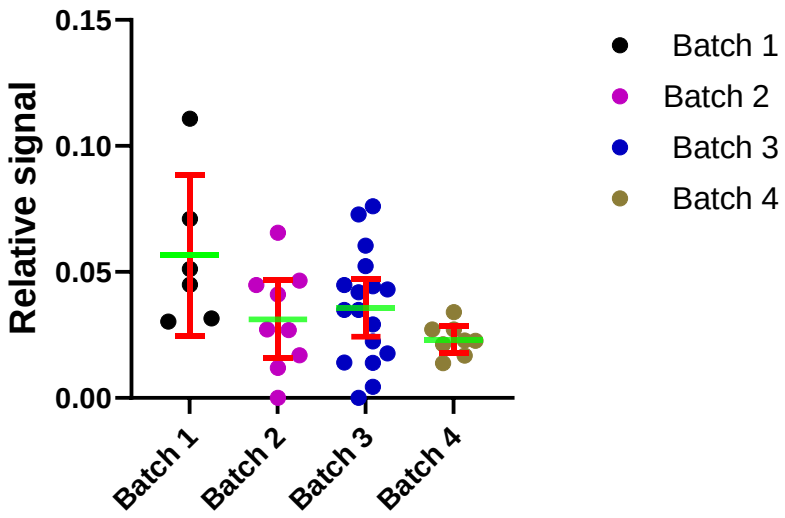

Figure 10: univariate scatter plot of different batches

Figure 10 is a column scatter plot of different batches showing significant differences in their means according to how they were processed.

Sample run order was not fully and properly randomized. Specifically, the first batch was not randomized prior to processing while the rest were randomized. Due to this randomization error, there was concern to know if a batch effect could be responsible for the significant difference in mean between COVID-19 positive and negative patient samples as depicted by the student t-test above. To make sure the batch effect was not the reason for the difference in means, a the Kruskal-Wallis test was performed on the data grouped by batch (Fig. 6), and the overall results showed no significant batch effect ($p > 0.05$).

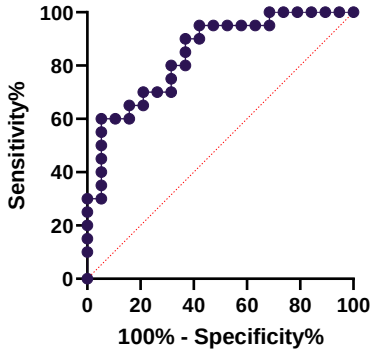

Figure 11: ROC plot for 3,6-GalNAc

Figure 11 is a ROC curve for 3,6 GalNAc showing possible connection/trade-offs between clinical sensitivity and specificity from the univariant scatter plot in Figure 5 for the Mayo urine Covid-19 Positive and Negative individuals.

The receiver operating characteristics (ROC) is a graph that shows the diagnostic capability of a classification model. The ROC curve is used to obtain the overall clinical performance of a diagnostic test. A ROC Plot was made for the 3,6-GalNAc node to indicate the trade-off between clinical sensitivity and specificity of data points from the univariant scatter plot.

The graph shows that the curve trends along the upper left side of the plot, and the area under the curve (AUC), 0.84, confirms the potential utility of the 3,6-GalNAc glycan node as a urinary biomarker for Covid-19 disease.

Discussion:

Potential clinical meaning of 3,6-GalNAc node as a urinary Covid-19 biomarker**.**

The 3,6 GalNAc node arises predominantly from O-linked glycans with branched core structures. Mucins are a class of glycoproteins that house an enriched number of these O-GalNAcs glycans known as the mucin-type-O-Glycans. In the human body, mucins line epithelial cell surfaces such as the gastrointestinal tract (GI), the urinogenital system (urinary and reproductive), and the respiratory tract. Since mucins line the urinary system, there is a high probability of finding O-GalNAc glycans, including 3,6-GalNAc in urine. However, the relative abundance of the 3,6-GalNAc node in the Covid-19 positive individuals was lower when compared to the Covid-19 negative individuals. This can be explained by assuming that there is a default number-of-mucins per glycoprotein ratio carrying the O-GalNAc glycans lining the urinogenital system of Covid-19 negative individuals. In a situation whereby we have improper kidney filtration (Covid-19 positive patients), additional non-mucinous glycoproteins may become more abundant in the urine than normal—decreasing the number-of-mucins per glycoprotein ratio relative to what is normally present. Meaning, the concentration of mucin per glycoprotein ratio is diminished when compared to the Covid-19 individuals where nothing is being abnormally added to the urine. As such, the 3,6-GalNAc node is expected to be lower in the Covid-19 positive patients. So, I hypothesized that kidney-damaged cells, caused by

the Covid-19 virus, SARS-CoV-2 might be responsible for the improper kidney filtration, hence the decrease of 3,6-GalNAc node in the Covid-19 positive patients.

CONCLUSION:

In this study, bottom-up glycan node analysis approach was employed to identify a urinary glycan-based marker of COVID-19 known as 3,6-GalNAc. 3,6-GalNAc, a uniquely linked monosaccharide derived from mostly core-branched O-glycans, was lower in the Covid-19 positive patients than in the negative individuals. Clinical validation of this marker will need to take place in the future, along with an investigation into the pathophysiology responsible for its decreased concentration in the urine of COVID-19 positive patients.

REFERENCES

1. Elflein, J. (2021). Coronavirus (COVID-19) in the US-statistics & facts. Statista.

2. Pisitkun, T., Johnstone, R., & Knepper, M. A. (2006). Discovery of urinary biomarkers. *Molecular & Cellular Proteomics*, *5*(10), 1760-1771.

3. Menez, S., Moledina, D. G., Thiessen-Philbrook, H., Wilson, F. P., Obeid, W., Simonov, M., ... & Nadkarni, G. (2022). Prognostic significance of urinary biomarkers in patients hospitalized with COVID-19. *American Journal of Kidney Diseases*, *79*(2), 257-267.

4. Mussap, M., Zaffanello, M., & Fanos, V. (2018). Metabolomics: a challenge for detecting and monitoring inborn errors of metabolism. *Annals of translational medicine*, *6*(17).

5. Rini, J. M., Moremen, K. W., Davis, B. G., & Esko, J. D. (2022). Glycosyltransferases and glycan-processing enzymes. *Essentials of Glycobiology [Internet]. 4th edition.*

6. Shajahan, A., Supekar, N. T., Gleinich, A. S., & Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, *30*(12), 981-988.

7. Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B., & Dwek, R. A. (1996). A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Analytical biochemistry*, *240*(2), 210-226.

8. Pabst, M., & Altmann, F. (2011). Glycan analysis by modern instrumental methods. *Proteomics*, *11*(4), 631-643.

9. Rothman, R. J., & Warren, L. (1988). Analysis of IgG glycopeptides by alkaline borate gel filtration chromatography. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *955*(2), 143-153.

10. Merkle, R. K., & Cummings, R. D. (1987). [18] Lectin affinity chromatography of glycopeptides. In Methods in enzymology (Vol. 138, pp. 232-259). Academic Press.

11. Zaare, S., Aguilar, J. S., Hu, Y., Ferdosi, S., & Borges, C. R. (2016). Glycan node analysis: a bottom-up approach to glycomics. *JoVE (Journal of Visualized Experiments)*, (111), e53961.

12. Walker, S. A., Aguilar Díaz De león, J. S., Busatto, S., Wurtz, G. A., Zubair, A. C., Borges, C. R., & Wolfram, J. (2020). Glycan node analysis of plasma-derived extracellular vesicles. *Cells*, *9*(9), 1946.

13. Borges, C. R., Rehder, D. S., & Boffetta, P. (2013). Multiplexed surrogate analysis of glycosyltransferase activity in whole biospecimens. *Analytical chemistry*, *85*(5), 2927-2936.

14. Heiss, C., Klutts, J. S., Wang, Z., Doering, T. L., & Azadi, P. (2009). The structure of Cryptococcus neoformans galactoxylomannan contains β-d-glucuronic acid. *Carbohydrate Research*, *344*(7), 915-920.

15. Xu, M., Yang, A., Xia, J., Jiang, J., Liu, C. F., Ye, Z., ... & Yang, S. (2022). Protein glycosylation in urine as a biomarker of diseases. *Translational Research*.

16. George, S., Pal, A. C., Gagnon, J., Timalsina, S., Singh, P., Vydyam, P., ... & Yale IMPACT Team. (2021). Evidence for SARS-CoV-2 spike protein in the urine of COVID-19 patients. Kidney360, 2(6), 924.

17. García-Ayllón, M. S., Moreno-Pérez, O., García-Arriaza, J., Ramos-Rincón, J. M., Cortés-Gómez, M. Á., Brinkmalm, G., ... & Sáez-Valero, J. (2021). Plasma ACE2 species are differentially altered in COVID-19 patients. The FASEB Journal, 35(8).

18. Hu, Y., Ferdosi, S., Kapuruge, E. P., Diaz de Leon, J. A., Stücker, I., Radoï, L., ... & Borges, C. R. (2019). Diagnostic and prognostic performance of blood plasma glycan features in the women epidemiology lung cancer (WELCA) Study. Journal of proteome research, 18(11), 3985-3998.

19. https://glygen.ccrc.uga.edu/ccrc/specdb/ms/pmaa/pframe.html

20. Romagnani, P., Remuzzi, G., Glassock, R., Levin, A., Jager, K. J., Tonelli, M., ... & Anders, H. J. (2017). Chronic kidney disease. *Nature reviews Disease primers*, *3*(1), 1-24.