

# Identifying Sialylation Linkages at the Glycopeptide Level by Glycosyltransferase Labeling Assisted Mass Spectrometry (GLAMS)

He Zhu<sup>†</sup>, Shuaishuai Wang<sup>†</sup>, Ding Liu<sup>†</sup>, Lang Ding<sup>†</sup>, Congcong Chen<sup>†</sup>, Yunpeng Liu<sup>†</sup>, Zhigang  
Wu<sup>†,‡</sup>, Roni Bollag<sup>‡</sup>, Kebin Liu<sup>§</sup>, Jun Yin<sup>†</sup>, Cheng Ma<sup>†</sup>, Lei Li<sup>\*†</sup>, Peng George Wang<sup>\*†,¶</sup>

<sup>†</sup> Department of Chemistry, Georgia State University, Atlanta, GA 30303, United States

<sup>‡</sup> Georgia Cancer Center, Augusta University, Augusta, GA 30912, United States

<sup>§</sup> Department of Biochemistry and Molecular Biology, Augusta University, Augusta, GA 30912, United States

<sup>#</sup> Current address: College of Bioscience and Bioengineering, Hebei University of Science and Technology,  
Shijiazhuang, Hebei 050018, P. R. China

<sup>¶</sup> Current address: Department of Chemistry, Southern University of Science and Technology, Shenzhen,  
Guangdong 518055, P. R. China.

## ABSTRACT

Precise assignment of sialylation linkages at the glycopeptide level is of importance in  
bottom-up glycoproteomics, and is also an indispensable step to understand the function of  
glycoproteins in pathogen-host interactions and cancer progression. Even though some efforts  
have been dedicated to the discrimination of  $\alpha$ 2,3/ $\alpha$ 2,6-sialylated isomers, unambiguous  
identification of sialoglycopeptide isomers is still needed. Herein, an innovative strategy of  
glycosyltransferase labeling assisted mass spectrometry (GLAMS) was developed. After specific  
enzymatic labeling, oxonium ions from higher-energy C-trap dissociation (HCD) fragmentation  
of  $\alpha$ 2,3-sialoglycopeptides generate unique reporters to distinctly differentiate those of  $\alpha$ 2,6-

sialoglycopeptide isomers. Using this strategy, a total of 1,236 linkage-specific sialoglycopeptides were successfully identified from 161 glycoproteins in human serum.

## INTRODUCTION

Glycosylation is one of the most important post-translational modifications (PTMs) of proteins<sup>1</sup>. It plays an indispensable role in a wide variety of biological and physiological processes including protein folding<sup>2</sup>, cellular communication and signaling<sup>3</sup>, immune response<sup>4</sup>, and pathogen-host interactions<sup>5</sup>. It was estimated that more than half of proteins in the human proteome are glycosylated<sup>6</sup>. Glycosylation is generally classified into N-glycosylation and mucin-type O-glycosylation. In N-glycosylation, the glycans are attached to asparagine residues of proteins in the consensus motif Asn-X-Ser/Thr/Cys (X can be any amino acid except for proline)<sup>7</sup>, whereas the glycan moieties are linked to serine or threonine residues of proteins in O-glycosylation<sup>1</sup>. N-acetylneuraminic acid (Neu5Ac) is most common form of diverse sialic acids found on glycan chains in humans<sup>8</sup>. Neu5Ac is the frequently the terminal monosaccharide residue that linked to galactose (Gal) via an  $\alpha$ 2,3 or an  $\alpha$ 2,6 linkage<sup>9</sup>. The two Neu5Ac linkage isomers have important roles in pathogen binding. For example, avian influenza A viruses preferentially bind to respiratory epithelial cells that contain  $\alpha$ 2,3-linked Neu5Ac moieties, while human influenza viruses selectively bind to  $\alpha$ 2,6-linked Neu5Ac moieties<sup>10</sup>. In addition, altered sialylation linkages is closely correlated with health status. In patients diagnosed with prostate cancer, the structure of sialylated glycans released from the prostate-specific antigen (PSA) is rich in  $\alpha$ 2,3 linkages, whereas the structure of sialylated glycans from healthy people's PSA is predominantly via  $\alpha$ 2,6 linkages<sup>11</sup>. Without detailed information on sialylation linkages, it is challenging to decipher the biological functions of sialoglycoproteins.

A variety of methods, including sialidase treatment<sup>12</sup>, chemical derivatization<sup>13-21</sup>, and mass spectrometry (MS)<sup>22-24</sup>, have been used to distinguish  $\alpha$ 2,3 and  $\alpha$ 2,6-sialylated glycans, but the location of corresponding glycosylation sites is not evident. The emerging technology of MS-based glycoproteomics<sup>25-29</sup> provides a high-throughput approach to analyze site-specific protein glycosylation. Consequently, strategies for identifying sialylation linkages at the glycopeptide level are required. Established chemical approaches<sup>30, 31</sup> modify not only sialic acids of the glycan moiety but also acidic amino acid residues and the C-terminus of the peptide backbone, which increases the complexity of spectra analysis, especially in relation to large scale glycoproteomics data<sup>32</sup>. Pseudo MS<sup>3</sup> fragmentation<sup>22</sup> and ion mobility-mass spectrometry (IM-MS)<sup>33</sup> have also been applied to distinguish sialoglycopeptide isomers but this technology has not been **demonstrated** (?effective) for large-scale glycopeptides analysis. Even though the fragmentation of high energy C-trap dissociation (HCD) can achieve the high-throughput assignment of  $\alpha$ 2,3/ $\alpha$ 2,6-sialoglycopeptide isomers, there are still many ambiguous glycopeptides without sialylation linkage information<sup>34</sup>.

Enzymatic labeling has become a promising tool for probing proteins with specific glycosylation because of its high specificity and mild reaction condition<sup>35-41</sup>. In our previous research, a  $\beta$ 1,4-N-acetylgalactosaminyltransferase from *Campylobacter jejuni* (CgtA) was demonstrated to specifically recognize the Neu5Ac( $\alpha$ 2,3)Gal epitope and transfer N-acetylgalactosamine (GalNAc) or N-azidoacetylgalactosamine (GalNAz) to Gal residue via  $\beta$ 1,4 linkage<sup>42</sup>. Taking advantage of the specificity of CgtA, we herein report a new strategy for identification of  $\alpha$ 2,3 and  $\alpha$ 2,6 sialoglycopeptide isomers by glycosyltransferase labeling assisted mass spectrometry (GLAMS). As shown in **Scheme 1**, sialoglycopeptides (monosialylation for example) without linkages information were treated with GalNAz labeling by CgtA. After HCD

fragmentation, if only the oxonium ions of HexNAc (N-acetylhexosamine, m/z 204) and Neu5Ac (m/z 292) are observed in the spectrum, this identifies an  $\alpha$ 2,6 linkage; if one more oxonium ion m/z 245 from GalNAz is included, then it represents an  $\alpha$ 2,3 linkage. We demonstrated the capacity of GLAMS to effectively differentiate  $\alpha$ 2,3/ $\alpha$ 2,6 isomers using synthetic sialoglycopeptides, and applied it to profile sialylated N-glycopeptides digested from serum samples, in which we identified 1,236 linkage specific sialoglycopeptides with an improved version of pGlyco 2.0 software<sup>27</sup>.

## EXPERIMENTAL SECTION

### Chemicals and materials

Amicon Ultra-0.5 Centrifugal Filter Unit with NMWL of 30 kD was obtained from Millipore Sigma (St. Louis, MO). iSPE<sup>®</sup> HILIC SPE Cartridge was purchased from The Nest Group (Southborough, MA). XAmide medium was provided by Dr. Xinmiao Liang's lab (Dalian Institute of Chemical Physics, Chinese Academy of Sciences). Sequencing grade porcine trypsin was purchased from Promega (Madison, WI); Deionized water was produced by a Milli-Q A10 system from Millipore (Bedford, MA). Optimal LC-MS grade quality acetonitrile (ACN), formic acid (FA) and water were purchased from Thermo Fisher Scientific, Inc. (Waltham, MA). Dithiothreitol (DTT), 2-Iodoacetamide (IAA) and trifluoroacetic acid (TFA) were bought from Alfa Aesar (Ward Hill, MA). Ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>) was purchased from Millipore Sigma (St. Louis, MO). All other chemicals of analytical grade were commercially available.

Sialylated oligosaccharide standards including 3'-sialyllactose (3SL) and 6'-sialyllactose (6SL) were enzymatically synthesized using *Pasteurella multocida*  $\alpha$ 2,3-sialyltransferase 1 and *Photobacterium damsela*  $\alpha$ 2,6-sialyltransferase as described before<sup>43</sup>.

Serum samples from four healthy people, ten patients with colon cancer and four patients with liver cancer were provided by Georgia Cancer Center at Augusta University and stored at -80 °C until use. The protocol for serum sample preparation was approved by the Institutional Review Board of Augusta University and was performed in accordance with the Helsinki Declaration. All participants gave written informed consents.

## Protein digestion

The protein digestion was performed using a filter aided sample preparation (FASP)<sup>44</sup> method with minor modification. Approximately 1 mg of total proteins from human serum was subjected to FASP procedures. Samples were denatured in 20  $\mu$ L of lysis buffer (4% SDS, 100 mM DTT, 20 mM Tris-HCl, pH 7.6) at 95 °C for 10 min. When the solution was cooled down to room temperature, 1 mL of UA solution (8 M urea, 20 mM Tris-HCl, pH 8.5) was added. The solution was gradually transferred into a 30 kD filter and centrifugated at 15,000 rpm at 20 °C for 15 min. Then 100  $\mu$ L of IAA solution (50 mM IAA in UA solution) was added to the filter and incubated for 30 min in the darkness followed by centrifugation at 15,000 rpm for 10 min. After that, 100  $\mu$ L of UA solution and 200  $\mu$ L of digestion buffer (100  $\mu$ L of 50 mM  $\text{NH}_4\text{HCO}_3$  each time) were added and centrifuged at 15,000 rpm for 10 min, respectively. Finally, 20  $\mu$ g of trypsin was added into the filter and incubated at 37 °C for 12 h. The tryptic peptides were eluted using 100  $\mu$ L of 50 mM  $\text{NH}_4\text{HCO}_3$  for six times. The concentration of the collected peptides was measured based on absorbance at 280 nm with Nanodrop. Five hundred  $\mu$ g of the peptides were

taken and subsequently dried in Vacufuge at room temperature, then were lyophilized twice to remove  $\text{NH}_4\text{HCO}_3$ . The dried samples were stored at  $-20^\circ\text{C}$  for further use.

# **Enrichment of glycopeptides**

The enrichment of glycopeptides was performed using an iSPE HILIC cartridge by the procedure of ZIC-HILIC<sup>45</sup>. Briefly, the HILIC cartridge was pre-washed with 300  $\mu\text{L}$  of 0.1% TFA and conditioning with 600  $\mu\text{L}$  of 80% ACN containing 0.1% TFA. Peptide samples dissolved in 400  $\mu\text{L}$  of 80% ACN containing 0.1% TFA were loaded onto the cartridge. The flow-through was reloaded onto the column for another two times. Then the column was washed with 1.2 mL of 80% ACN containing 1% TFA. The glycopeptides were eluted using 750  $\mu\text{L}$  of 0.1% TFA, 60  $\mu\text{L}$  of  $\text{H}_2\text{O}$ , 60  $\mu\text{L}$  of 25 mM  $\text{NH}_4\text{HCO}_3$  and 60  $\mu\text{L}$  of 50% ACN. The fractions were combined followed by lyophilization and stored at  $-20^\circ\text{C}$ .

# **Enzymatic labeling of glycopeptides**

The  $\beta$ 1,4-N-acetylgalactosaminyltransferase from *C. jejuni* (CgtA) and sugar nucleotide UDP-N-azidoacetylgalactosamine (UDP-GalNAz) were prepared as described previously<sup>42</sup>. For enzymatic labeling, the 40  $\mu\text{L}$  reaction mixture containing the enriched glycopeptides, 50 mM Tris-HCl buffer (pH 7.5), 5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  UDP-GalNAz and 10  $\mu\text{g}$  CgtA was incubated in a  $37^\circ\text{C}$  water bath for 1 h. The reaction was quenched by adding 60  $\mu\text{L}$  of ACN, 6  $\mu\text{L}$  of 10% TFA and 215  $\mu\text{L}$  of ACN containing 1% TFA. Then the glycopeptides were purified with an in-house packed Amide HILIC microcolumn (6 mg of XAmide HILIC medium in a 200  $\mu\text{L}$  tip) which was preconditioned with 100  $\mu\text{L}$  of 0.1% TFA and 200  $\mu\text{L}$  of 85% ACN containing 0.1% TFA. After loading the reaction mixture onto the tip column, the column was washed with 100  $\mu\text{L}$  of 85% ACN, 0.1% TFA to remove salts and proteins. The glycopeptides were collected using 200  $\mu\text{L}$  of 0.1% TFA and lyophilized.

## Nano LC-MS/MS analysis

Nano RP HPLC-MS experiments were performed on an LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher) equipped with EASY-spray source and nano-LC UltiMate 3000 high-performance liquid chromatography system (Thermo Fisher). The separation was performed on an EASY-Spray PepMap C18 Column (75  $\mu\text{m} \times 50\text{ cm}$ , 2  $\mu\text{m}$ , ThermoFisher, US) with a linear gradient from 3% to 40% mobile phase B for 120 min at a flow rate of 300 nL/min (mobile phase A: 1.95% ACN, 97.95% H<sub>2</sub>O, 0.1% FA; mobile phase B: 79.95% ACN, 19.95% H<sub>2</sub>O, 0.1% FA). The mass spectrometer was operated in the data-dependent mode. A full-scan survey MS experiment (m/z range from 375 to 2000; automatic gain control target, 1,000,000 ions; resolution at 400 m/z, 60,000; maximum ion accumulation time, 50 ms) was acquired by the Orbitrap mass spectrometer, and ten most intense ions were fragmented by higher-energy C-trap dissociation (HCD) with an optimized stepped collision energy (SCE) of  $30 \pm 15\%$  NCE.

## Data analysis

The pGlyco 3.0 software (an improved version of pGlyco 2.0<sup>27</sup>) was used to identify N-glycopeptides. In the glycan database, we defined an unnatural monosaccharide GalNAz (Gaz, Z). We created all the possible “Gaz” containing human N-glycan sequences following the motif H(A)(Z), where “H” represents Galactose, “A” represents Neu5Ac and “Z” represents GalNAz. For the sialoglycopeptides identification, the raw data was first converted to an mgf file using pParse<sup>46</sup>. The potential sialoglycopeptides spectra containing oxonium ions of both GlcNAc (m/z 204.09, m/z 186.08) and Neu5Ac (m/z 292.10, m/z 274.09) were extracted and merged into a new mgf file using an in-house program. The new mgf file was searched against a Uniprot-Swiss human protein database (20,417 reviewed entries on March 20, 2019) and a modified N-glycan database. Other searching parameters were used including carbamidomethyl (Cys) as a

fixed modification, deamination (Asn) and oxidation (Met) as variable modifications, and trypsin as the digestion enzyme with maximum two missed cleavage sites allowed. The mass tolerance for the precursor ions and the fragment ions was set to 10 ppm and 20 ppm, respectively. A total false discovery rate (FDR) of 1% was applied to all data sets at the glycopeptide level. The identified glycoproteins were annotated using Uniprot ID mapping<sup>47</sup>.

## RESULTS AND DISCUSSION

### The specificity of CgtA for Sialoglycopeptides

Our previous study showed that CgtA had a high specificity for (Neu5Ac- $\alpha$ 2,3-Gal)-containing glycans<sup>42</sup>. To test its selectivity for sialoglycopeptides, we synthesized 23SP and 26SP (synthesis and data not published) as shown in **Figure 1**. The activity of CgtA was validated with sialolactose prior to the enzymatic reaction of sialoglycopeptides. After labeling with GalNAz by CgtA, only  $\alpha$ 2,3-sialylated lactose showed a mass shift of 244 Da in the MS spectrum and no substrate was obviously observed (**Figure S1**), which indicated that CgtA had good catalytic activity and high specificity as well.

ESI-MS was used to monitor the GalNAz labeling reaction for sialoglycopeptides. For the  $\alpha$ 2,3-sialoglycopeptide 23SP, the monoisotopic peak at  $m/z$  1032.0808 is from the triply charged glycopeptide (**Figure S2A**). After incubating with CgtA and UDP-GalNAz for 30 min, a new dominant peak at  $m/z$  1113.4422 with three positive charges was observed, which correspond to the GalNAz labeled 23SP (**Figure S2C**). However, we could still detect minor substrate 23SP. The  $\alpha$ 2,6-sialoglycopeptide 26SP has the same exact mass (**Figure S2B**) as 23SP. After GalNAz labeling by CgtA, the strongest peak is located at the same exact position, and a



trace amount of product at  $m/z$  1113.4422 was detected (**Figure S2D**). This may result from the very weak activity of CgtA for 2,6-sialylated glycans<sup>42</sup>.

The reaction time was further optimized to increase the conversion rate of GalNAz labeling of 23SP. The reaction mixture at different time points was analyzed by LC-MS and the reaction efficiency was calculated based on the peak area. For 23SP, in the first half an hour, the reaction efficiency was dramatically increased. After that, the increase slowed down and the reaction efficiency approached 100% at 60 min. Meanwhile, the conversion rate for the 26SP remained unchanged (less than 1%) over time (**Figure 2**). Finally, we chose 60 min as the reaction time for the subsequent GalNAz labeling reactions.

## Identifying Sialoglycopeptide Linkage Isomers by GalNAz Labeling and Tandem MS

Although discrimination of  $\alpha 2,3/\alpha 2,6$ -sialoglycopeptide isomers can be achieved by the mass shift in primary MS as shown above, we cannot accurately determine the glycopeptide sequence using only molecular weight information. Tandem MS is an effective way for sequencing glycopeptides, and HCD-MS/MS is commonly used to analyze glycopeptides because of the high accuracy and detectable diagnostic oxonium ions of monosaccharides<sup>26-28, 48</sup>. It was also demonstrated that HCD with stepped collision energy (SCE) can simultaneously determine the sequences of both glycans and peptides in a single run<sup>27, 48</sup>. To achieve a better sequence coverage of glycopeptides on the Orbitrap we used, SCE was optimized with glycopeptides enriched from human serum. The identification of N-glycopeptides was performed using pGlyco 2.0 software<sup>27</sup>. A series of two-step SCEs and three-step SCEs were employed to compare the number of identified glycopeptides. As shown in **Figure S3**, with a three-step SCE of “15-30-45”, more glycopeptides can be identified from the same sample.

The optimized SCE-HCD-MS/MS was applied for sequencing 23SP and 26SP. After GalNAz labeling, 23SP (**Figure 3A**) showed diagnostic oxonium ions of GlcNAc and Neu5Ac at  $m/z$  204.09 and 292.10, which are common features of sialoglycopeptides. Apart from these oxonium ions, 23SP also had a strong unknown peak at  $m/z$  245.09. Compared with the spectrum of 23SP before GalNAz labeling (**Figure S4**), we can obviously see that the peak with  $m/z$  245.09 corresponds to the oxonium ion of GalNAz. Following the dissociation pathway of GlcNAc<sup>39, 49, 50</sup> in **Scheme S1A**, a series of fragmentation ions of GalNAz in HCD were illustrated in **Scheme S1B**. The peaks at  $m/z$  245.09,  $m/z$  227.08,  $m/z$  209.07 and  $m/z$  179.06, which match the theoretical values very well, are diagnostic ions of GalNAz (**Figure S5**). For 26SP (**Figure 3B**), its spectrum showed that it shared similar fragmentation patterns with GalNAz labeled 23SP. In the low  $m/z$  region, only diagnostic ions of GlcNAc and Neu5Ac were observed, which clearly demonstrated that the GLAMS is an effective approach to distinguish intact linkage specific sialoglycopeptides.

### Linkage Specific Sialoglycopeptides in Human Serum

We then applied this method to analyze the N-linked  $\alpha$ 2,3/ $\alpha$ 2,6-sialoglycopeptides of complex biological samples. Serum samples from healthy individuals ( $n = 4$ ) and patients diagnosed with liver cancer ( $n = 4$ ) and colon cancer ( $n = 10$ ) were used in this study. In the beginning, we tried to perform the enzymatic labeling of the mixture of tryptic peptides and glycopeptides. After drying from the FASP eluents, the mixture was not completely dissolved in the Tris buffer (pH 7.5) for the enzymatic reaction, which may result from the existence of some basic peptides. To solve the solubility issue, we adopted step-by-step glycopeptides enrichment and enzymatic labeling as shown in the workflow **Figure 4**. It turned out that the enriched

glycopeptides were dissolved in the basic buffer very well. This may be attributed to that conjugated glycans increase the hydrophilicity of peptides and some sialoglycans decrease the isoelectric points of peptides according to the previous findings<sup>51, 52</sup>.

All the serum samples were analyzed in three replicated LC-MS/MS runs for a large-scale study of N-linked sialoglycopeptides. Using the oxonium ions of GalNAz as a tag to mark  $\alpha$ 2,3-sialylation, a total of 22,819 sialoglycopeptide spectra were identified with a 1% total FDR, corresponding to 1,236 unique site-specific sialoglycans on 261 N-glycosylation sites from 161 unique glycoproteins (**Table S1**), among of which 53% was also identified previously using a deglycosylation method<sup>53</sup> as shown in **Figure S6**. For the distribution analysis of  $\alpha$ 2,3 or  $\alpha$ 2,6-sialylation (**Figure 5A**), only 48 distinct site-specific sialoglycopeptides contained pure  $\alpha$ 2,3-sialylation linkage, 236 sialoglycopeptides contained hybrid sialylation, and all the rest were uniquely 2,6-sialylated, which accounts for more than 77% of the identified sialoglycopeptides. This is consistent with the previous result that  $\alpha$ 2,6-sialylation is dominant in sialylated N-glycans released from human serum samples<sup>54</sup>. Taking a closer look at the distribution of sialylation levels, monosialylation is the major component in strictly  $\alpha$ 2,3 or  $\alpha$ 2,6-sialoglycopeptides, and 74% of the hybrid sialoglycopeptides are disialylated.

The percentage of the overall overlapped glycosites (**Figure 5C**) and glycoproteins (**Figure 5D**) in a specific type of serum is 72-80% and 65-78%, respectively. The percentage of overlapping sialoglycopeptides (**Figure 5B**) in each kind of serum sample was significantly lower than those of the glycosites and glycoproteins, reflecting the diversity of site-specific glycosylation, especially for samples from colon cancer and liver cancer patients. A highly glycosylated protein P01876 (IGHA1) was selected to demonstrate the diversity of site-specific glycosylation in the three different sera. The previous study showed that IgA was  $\alpha$ 2,3 and  $\alpha$ 2,6-

sialylated on N-glycans<sup>55</sup>. Here, we identified 35 unique sialoglycopeptides including  $\alpha$ 2,3 and  $\alpha$ 2,6 linkages on one N-glycosylation site of IGHA1 (**Table S2**). In addition, we found some unique sialoglycopeptides of IGHA1 expressed in a serum-specific manner. For example, only unique  $\alpha$ 2,6-sialoglycopeptides were identified in serum from liver cancer patients (representative spectrum in **Figure S7**) and normal people (representative spectrum in **Figure S8**), whereas unique  $\alpha$ 2,6-sialoglycopeptides and one hybrid sialoglycopeptide were detected from colon cancer patients' sera (representative spectra in **Figure S9**).

## CONCLUSIONS

Elucidating linkage-specific sialylation of glycoproteins is significant to understand pathogen-host interactions and disease development. By integrating enzymatic labeling and MS-based glycoproteomics, we developed a high-throughput glycosyltransferase labeling assisted mass spectrometry (GLAMS) to identify sialoglycopeptide linkage isomers. Facilitated by the acceptor specificity of the glycosyltransferase CgtA toward the Neu5Ac( $\alpha$ 2,3)Gal epitope, we have demonstrated the distinct diagnostic oxonium ions as tags to effectively discriminate  $\alpha$ 2,3- and  $\alpha$ 2,6-sialoglycopeptide isomers using LC separation coupled with SCE-HCD-MS/MS fragmentation. We also successfully applied this approach to a large-scale profiling of linkage-specific sialoglycopeptides in human serum samples. The strategy of GLAMS should be of great benefit in investigating differential changes in response to disease states from a glycoproteomics perspective, and has potential to contribute to biomarker discovery in many diseases.

## AUTHOR INFORMATION

### Corresponding Authors

[lli22@gsu.edu](mailto:lli22@gsu.edu) (L. L.) or [pwang11@gsu.edu](mailto:pwang11@gsu.edu) (P. G. W.)

### Author Contributions

H.Z., L. L., and P. G. W. conceived the project; H.Z. designed and performed the experiments, analyzed the data, and wrote the manuscript; S. W. synthesized the sialoglycopeptides; D. L. and L. D. prepared the serum glycopeptides for collision energy optimization; C. C., Y. L. and Z. W. synthesized the UDP-GalNAz; R. B. and K. L. collected and prepared the human serum samples; C. M., J. Y., L. L., and P. G. W. edited the manuscript. All the authors revised and approved the manuscript.

## ACKNOWLEDGEMENTS

We sincerely thank Dr. Wenfeng Zeng from the Key Lab of Intelligent Information Processing of Chinese Academy of Sciences for the help of parameter setting of pGlyco 3.0, Georgia Cancer Center Biorepository for providing human serum samples, and Georgia Research Alliance for financial support to purchase mass spectrometers. This work is supported by National Institute of Health (U01GM116263 and U54HL142019).

## CONFLICT OF INTERESTS

The authors declare no conflict of interest.

## 296 REFERENCES

- 297 1. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease.  
298 *Cell* **2006**, 126 (5), 855-867.
- 299 2. Banerjee, S.; Vishwanath, P.; Cui, J.; Kelleher, D. J.; Gilmore, R.; Robbins, P. W.;  
300 Samuelson, J., The evolution of N-glycan-dependent endoplasmic reticulum quality control  
301 factors for glycoprotein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104 (28),  
302 11676-11681.
- 303 3. Furukawa, K.; Ohkawa, Y.; Yamauchi, Y.; Hamamura, K.; Ohmi, Y.; Furukawa, K.,  
304 Fine tuning of cell signals by glycosylation. *J. Biochem.* **2012**, 151 (6), 573-578.
- 305 4. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and  
306 the immune system. *Science* **2001**, 291 (5512), 2370-2376.
- 307 5. Karlsson, K. A., Pathogen-host protein-carbohydrate interactions as the basis of  
308 important infections. *Molecular Immunology of Complex Carbohydrates-2* **2001**, 491, 431-443.
- 309 6. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as  
310 deduced from analysis of the SWISS-PROT database. *BBA-Gen. Subjects* **1999**, 1473 (1), 4-8.
- 311 7. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M., Precision Mapping of an In  
312 Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **2010**, 141 (5),  
313 897-907.
- 314 8. Varki, N. M.; Varki, A., Diversity in cell surface sialic acid presentations: implications  
315 for biology and disease. *Lab. Invest.* **2007**, 87 (9), 851-857.
- 316 9. Varki, A.; Schauer, R., Sialic Acids. In *Essentials of Glycobiology*, nd; Varki, A.;  
317 Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler,  
318 M. E., Eds. Cold Spring Harbor (NY), 2009.
- 319 10. Stevens, J.; Blixt, O.; Paulson, J. C.; Wilson, I. A., Glycan microarray technologies:  
320 tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **2006**, 4 (11), 857-864.
- 321 11. Hatakeyama, S.; Yoneyama, T.; Tobisawa, Y.; Ohyama, C., Recent progress and  
322 perspectives on prostate cancer biomarkers. *Int. J. Clin. Oncol.* **2017**, 22 (2), 214-221.
- 323 12. Mancera-Artu, M.; Gimenez, E.; Barbosa, J.; Sanz-Nebot, V., Identification and  
324 characterization of isomeric N-glycans of human alfa-acid-glycoprotein by stable isotope  
325 labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion. *Anal. Chim. Acta.*  
326 **2016**, 940, 92-103.
- 327 13. Mechref, Y.; Kang, P.; Novotny, M. V., Differentiating structural isomers of sialylated  
328 glycans by matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass  
329 spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, 20 (8), 1381-1389.
- 330 14. Alley, W. R., Jr.; Novotny, M. V., Glycomic analysis of sialic acid linkages in glycans  
331 derived from blood serum glycoproteins. *J. Proteome. Res.* **2010**, 9 (6), 3062-3072.
- 332 15. Jiang, K.; Zhu, H.; Li, L.; Guo, Y. X.; Gashash, E.; Ma, C.; Sun, X. L.; Li, J.; Zhang,  
333 L. W.; Wang, P. G., Sialic acid linkage-specific permethylation for improved profiling of protein  
334 glycosylation by MALDI-TOF MS. *Anal. Chim. Acta.* **2017**, 981, 53-61.
- 335 16. Wheeler, S. F.; Domann, P.; Harvey, D. J., Derivatization of sialic acids for stabilization  
336 in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation  
337 of  $\alpha(2\rightarrow3)$  and  $\alpha(2\rightarrow6)$ -isomers. *Rapid Commun. Mass Spectrom.* **2009**, 23 (2), 303-312.
- 338 17. Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wührer, M., High-  
339 Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-  
340 Specific Sialic Acid Esterification. *Anal. Chem.* **2014**, 86 (12), 5784-5793.

18. Holst, S.; Heijs, B.; de Haan, N.; van Zeijl, R. J. M.; Briare-de Bruijn, I. H.; van Pelt, G. W.; Mehta, A. S.; Angel, P. M.; Mesker, W. E.; Tollenaar, R. A.; Drake, R. R.; Bovee, J. V. M. G.; McDonnell, L. A.; Wuhler, M., Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal. Chem.* **2016**, *88* (11), 5904-5913.
19. Yang, S.; Jankowska, E.; Kosikova, M.; Xie, H.; Cipollo, J., Solid-Phase Chemical Modification for Sialic Acid Linkage Analysis: Application to Glycoproteins of Host Cells Used in Influenza Virus Propagation. *Anal. Chem.* **2017**, *89* (17), 9508-9517.
20. Li, H. H.; Gao, W. J.; Feng, X. J.; Liu, B. F.; Liu, X., MALDI-MS analysis of sialylated N-glycan linkage isomers using solid-phase two step derivatization method. *Anal. Chim. Acta.* **2016**, *924*, 77-85.
21. Nishikaze, T.; Tsumoto, H.; Sekiya, S.; Iwamoto, S.; Miura, Y.; Tanaka, K., Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Anal. Chem.* **2017**, *89* (4), 2353-2360.
22. Chen, C. H.; Lin, Y. P.; Lin, J. L.; Li, S. T.; Ren, C. T.; Wu, C. Y.; Chen, C. H., Rapid Identification of Terminal Sialic Acid Linkage Isomers by Pseudo-MS3 Mass Spectrometry. *Isr. J. Chem.* **2015**, *55* (3-4), 412-422.
23. Guttman, M.; Lee, K. K., Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal. Chem.* **2016**, *88* (10), 5212-5217.
24. Depland, A. D.; Renois-Predelus, G.; Schindler, B.; Compagnon, I., Identification of sialic acid linkage isomers in glycans using coupled InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and mass spectrometry. *Int. J. Mass Spectrom.* **2018**, *434*, 65-69.
25. Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R., Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* **2015**, *12* (6), 561-567.
26. Sun, S. S.; Shah, P.; Eshghi, S. T.; Yang, W. M.; Trikanad, N.; Yang, S.; Chen, L. J.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H., Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **2016**, *34* (1), 84-88.
27. Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y., pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun* **2017**, *8* (1), 438.
28. Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattlinger, A.; Durnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M., Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **2017**, *549* (7673), 538-542.
29. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat Commun* **2019**, *10* (1), 1311.
30. Nishikaze, T.; Nakamura, T.; Jinmei, H.; Amano, J., Negative-ion MALDI-MS2 for discrimination of alpha2,3- and alpha2,6-sialylation on glycopeptides labeled with a pyrene derivative. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2011**, *879* (17-18), 1419-1428.
31. de Haan, N.; Reiding, K. R.; Habegger, M.; Reusch, D.; Falck, D.; Wuhler, M., Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Anal. Chem.* **2015**, *87* (16), 8284-8291.



32. Yang, S.; Wu, W. W.; Shen, R. F.; Bern, M.; Cipollo, J., Identification of Sialic Acid Linkages on Intact Glycopeptides via Differential Chemical Modification Using IntactGIG-HILIC. *J. Am. Soc. Mass Spectrom.* **2018**, 29 (6), 1273-1283.
33. Hinneburg, H.; Hofmann, J.; Struwe, W. B.; Thader, A.; Altmann, F.; Silva, D. V.; Seeberger, P. H.; Pagel, K.; Kolarich, D., Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chem. Commun.* **2016**, 52 (23), 4381-4384.
34. Pett, C.; Nasir, W.; Sihlbom, C.; Olsson, B. M.; Caixeta, V.; Schorlemer, M.; Zahedi, R. P.; Larson, G.; Nilsson, J.; Westerlind, U., Effective Assignment of alpha 2,3/alpha 2,6-Sialic Acid Isomers by LC-MS/MS-Based Glycoproteomics. *Angew Chem. Int. Ed. Engl.* **2018**, 57 (30), 9320-9324.
35. Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.; Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C., A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.* **2003**, 125 (52), 16162-16163.
36. Chaubard, J. L.; Krishnamurthy, C.; Yi, W.; Smith, D. F.; Hsieh-Wilson, L. C., Chemoenzymatic probes for detecting and imaging fucose-alpha(1-2)-galactose glycan biomarkers. *J. Am. Chem. Soc.* **2012**, 134 (10), 4489-4492.
37. Mbua, N. E.; Li, X.; Flanagan-Steet, H. R.; Meng, L.; Aoki, K.; Moremen, K. W.; Wolfert, M. A.; Steet, R.; Boons, G. J., Selective exo-enzymatic labeling of N-glycans on the surface of living cells by recombinant ST6Gal I. *Angew Chem. Int. Ed. Engl.* **2013**, 52 (49), 13012-13015.
38. Li, Q.; Li, Z.; Duan, X.; Yi, W., A tandem enzymatic approach for detecting and imaging tumor-associated Thomsen-Friedenreich antigen disaccharide. *J. Am. Chem. Soc.* **2014**, 136 (36), 12536-12539.
39. Zheng, J.; Xiao, H.; Wu, R., Specific Identification of Glycoproteins Bearing the Tn Antigen in Human Cells. *Angew Chem. Int. Ed. Engl.* **2017**, 56 (25), 7107-7111.
40. Wen, L.; Liu, D.; Zheng, Y.; Huang, K.; Cao, X.; Song, J.; Wang, P. G., A One-Step Chemoenzymatic Labeling Strategy for Probing Sialylated Thomsen-Friedenreich Antigen. *ACS Cent. Sci.* **2018**, 4 (4), 451-457.
41. Wu, Z. L. L.; Person, A. D.; Anderson, M.; Burroughs, B.; Tatge, T.; Khatri, K.; Zou, Y. L.; Wang, L. C.; Geders, T.; Zaia, J.; Sackstein, R., Imaging specific cellular glycan structures using glycosyltransferases via click chemistry. *Glycobiology* **2018**, 28 (2), 69-79.
42. Wen, L. Q.; Zheng, Y.; Jiang, K.; Zhang, M. Z.; Kondengaden, S. M.; Li, S. S.; Huang, K.; Li, J.; Song, J.; Wang, P. G., Two-Step Chemoenzymatic Detection of N-Acetylneuraminic Acid-alpha(2-3)-Galactose Glycans. *J. Am. Chem. Soc.* **2016**, 138 (36), 11473-11476.
43. Wu, Z.; Liu, Y.; Ma, C.; Li, L.; Bai, J.; Byrd-Leotis, L.; Lasanajak, Y.; Guo, Y.; Wen, L.; Zhu, H.; Song, J.; Li, Y.; Steinhauer, D. A.; Smith, D. F.; Zhao, B.; Chen, X.; Guan, W.; Wang, P. G., Identification of the binding roles of terminal and internal glycan epitopes using enzymatically synthesized N-glycans containing tandem epitopes. *Org. Biomol. Chem.* **2016**, 14 (47), 11106-11116.
44. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, 6 (5), 359-362.



45. Parker, B. L.; Thaysen-Andersen, M.; Solis, N.; Scott, N. E.; Larsen, M. R.; Graham, M. E.; Packer, N. H.; Cordwell, S. J., Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J. Proteome. Res.* **2013**, *12* (12), 5791-5800.
46. Yuan, Z. F.; Liu, C.; Wang, H. P.; Sun, R. X.; Fu, Y.; Zhang, J. F.; Wang, L. H.; Chi, H.; Li, Y.; Xiu, L. Y.; Wang, W. P.; He, S. M., pParse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra. *Proteomics* **2012**, *12* (2), 226-235.
47. UniProt, C., The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2007**, *35* (Database issue), D193-197.
48. Pioch, M.; Hoffmann, M.; Pralow, A.; Reichl, U.; Rapp, E., glyXtool(MS): An Open-Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data. *Anal. Chem.* **2018**, *90* (20), 11908-11916.
49. Zhao, P.; Viner, R.; Teo, C. F.; Boons, G. J.; Horn, D.; Wells, L., Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome. Res.* **2011**, *10* (9), 4088-4104.
50. Halim, A.; Westerlind, U.; Pett, C.; Schorlemer, M.; Ruetschi, U.; Brinkmalm, G.; Sihlbom, C.; Lengqvist, J.; Larson, G.; Nilsson, J., Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome. Res.* **2014**, *13* (12), 6024-6032.
51. Willey, K. P., An elusive role for glycosylation in the structure and function of reproductive hormones. *Hum. Reprod. Update* **1999**, *5* (4), 330-355.
52. Hayakawa, S.; Nakai, S., Relationships of Hydrophobicity and Net Charge to the Solubility of Milk and Soy Proteins. *J. Food Sci.* **1985**, *50* (2), 486-491.
53. Zhu, H.; Li, X.; Qu, J.; Xiao, C.; Jiang, K.; Gashash, E.; Liu, D.; Song, J.; Cheng, J.; Ma, C.; Wang, P. G., Diethylaminoethyl Sepharose (DEAE-Sepharose) microcolumn for enrichment of glycopeptides. *Anal. Bioanal. Chem.* **2017**, *409* (2), 511-518.
54. Jiang, K.; Zhu, H.; Li, L.; Guo, Y.; Gashash, E.; Ma, C.; Sun, X.; Li, J.; Zhang, L.; Wang, P. G., Sialic acid linkage-specific permethylation for improved profiling of protein glycosylation by MALDI-TOF MS. *Anal. Chim. Acta.* **2017**, *981*, 53-61.
55. Maurer, M. A.; Meyer, L.; Bianchi, M.; Turner, H. L.; Le, N. P. L.; Steck, M.; Wyrzucki, A.; Orłowski, V.; Ward, A. B.; Crispin, M.; Hangartner, L., Glycosylation of Human IgA Directly Inhibits Influenza A and Other Sialic-Acid-Binding Viruses. *Cell Rep.* **2018**, *23* (1), 90-99.

## Captions for Figures, and Scheme

**Scheme 1.** Strategy for distinguishing  $\alpha 2,3/\alpha 2,6$ -sialoglycopeptides by glycosyltransferase labeling assisted mass spectrometry.

**Figure 1.** Structure of the synthesized  $\alpha 2,3$ -sialoglycopeptides (23SP) and  $\alpha 2,6$ -sialoglycopeptides (26SP).

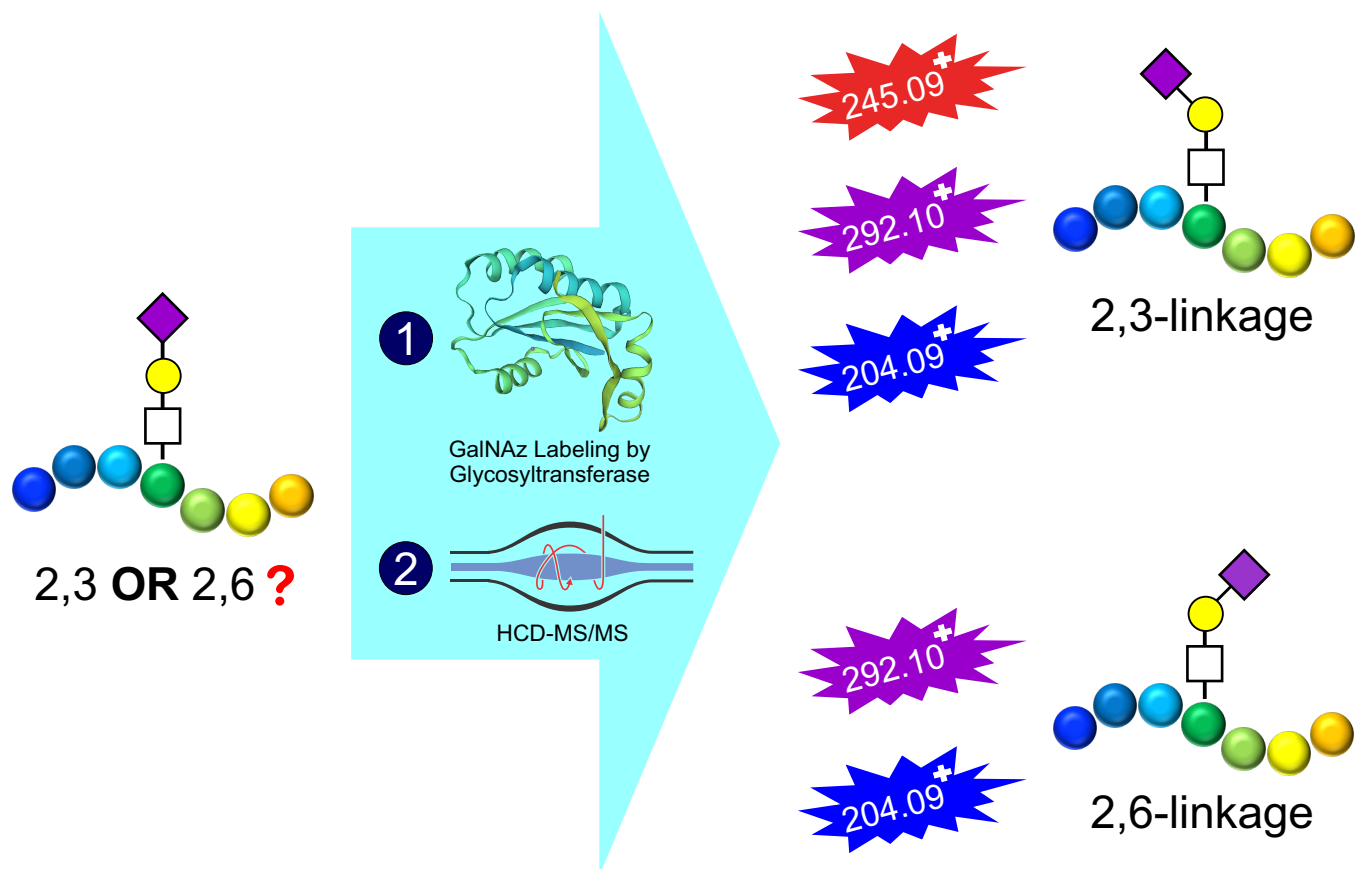
**Figure 2.** Time optimization for CgtA labeling of 23SP and 26SP.

**Figure 3.** HCD-MS/MS spectra of 23SP (A) and 26SP (B) after enzymatic labeling.

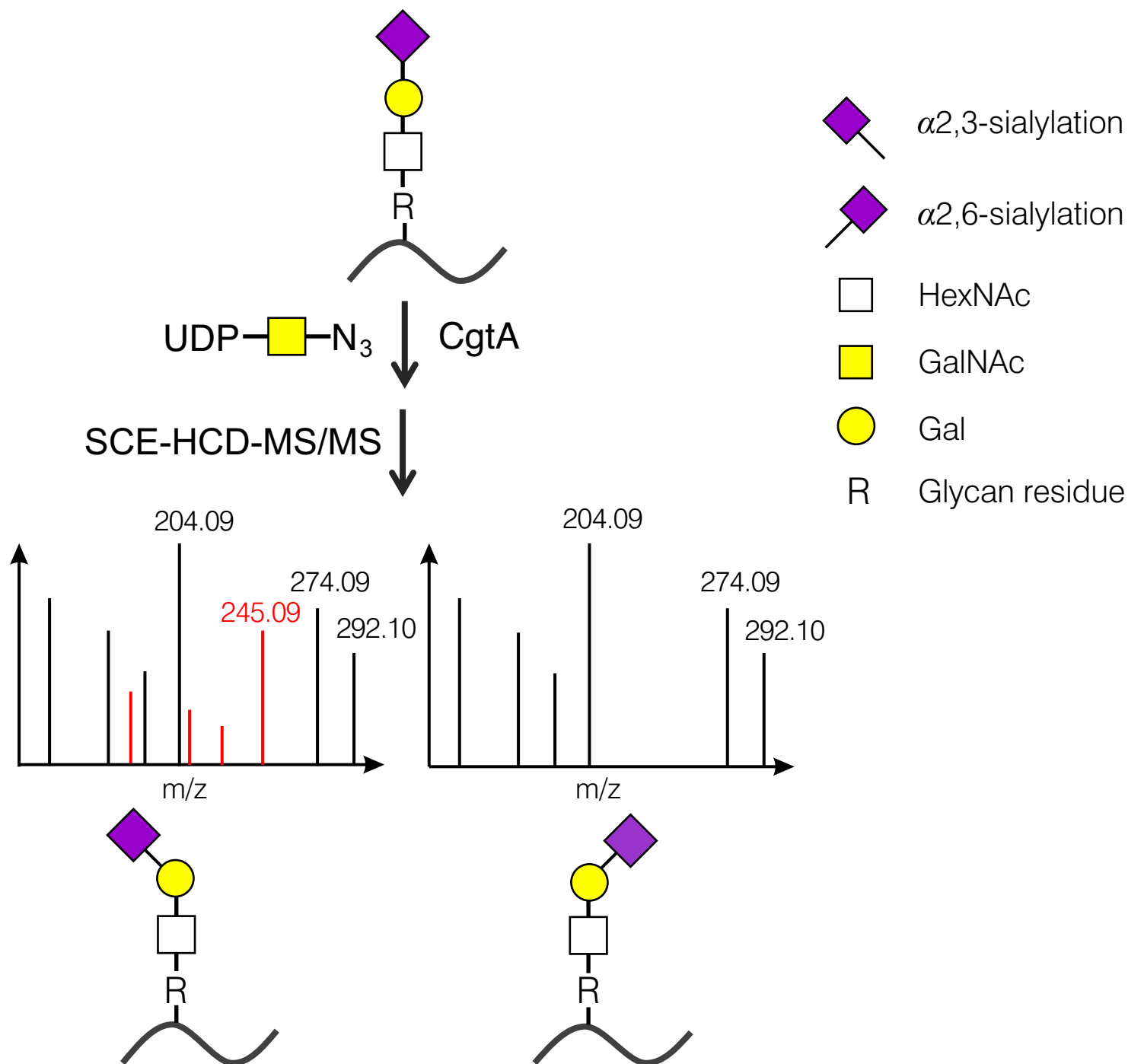
**Figure 4.** Workflow for identification of linkage specific sialoglycopeptides from human serum.

**Figure 5.** (A) Distribution of the sialoglycopeptides from human serum; Venn diagrams of identified glycopeptides (B), N-glycosylation sites (C), and glycoproteins (D) in serum samples from colon cancer patients, liver cancer patients and normal people.

# Abstract



# Scheme 1



## Figure 1

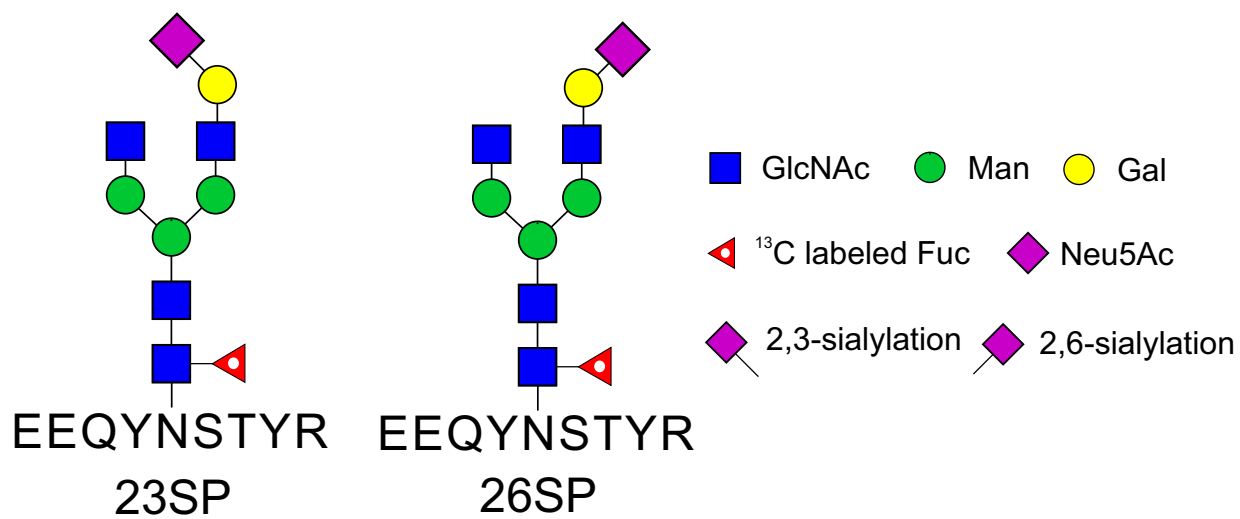
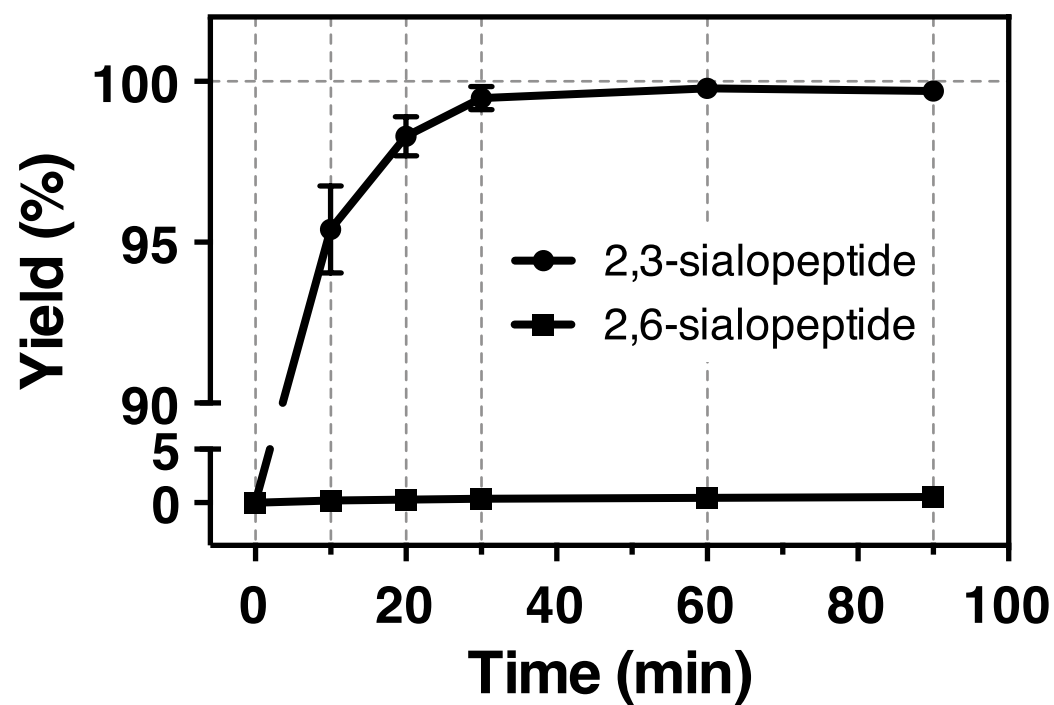
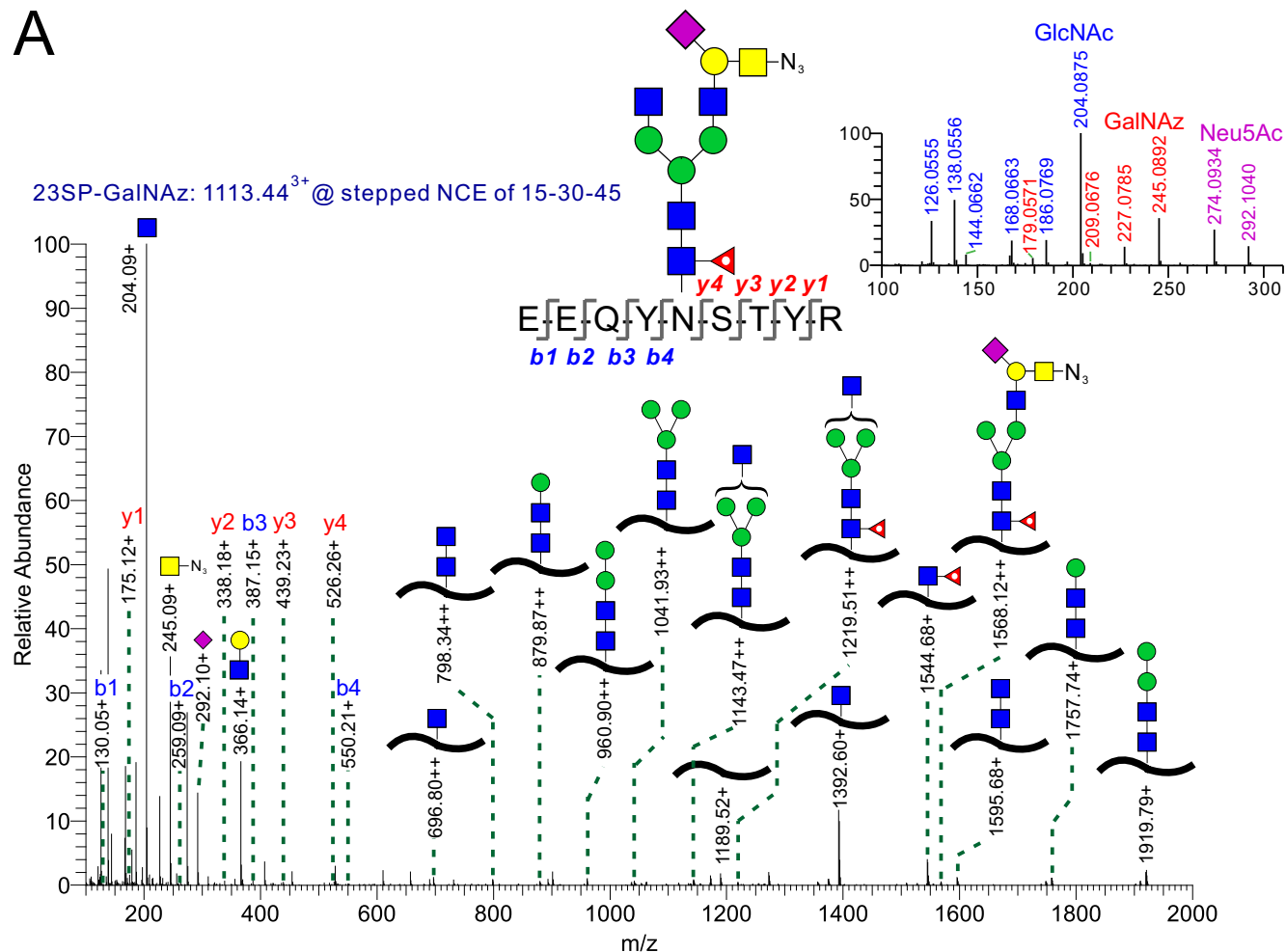


Figure 2

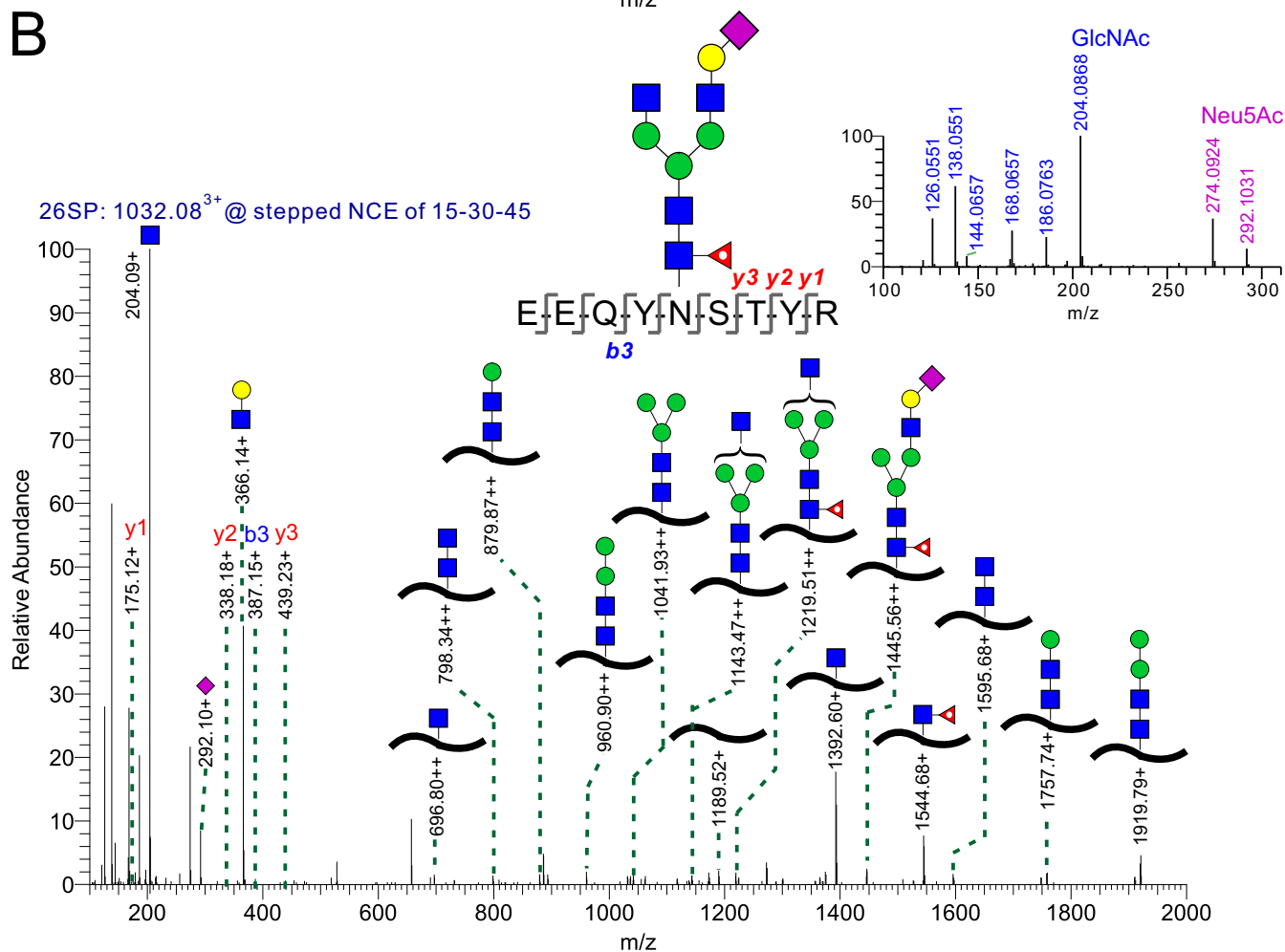


# Figure 3

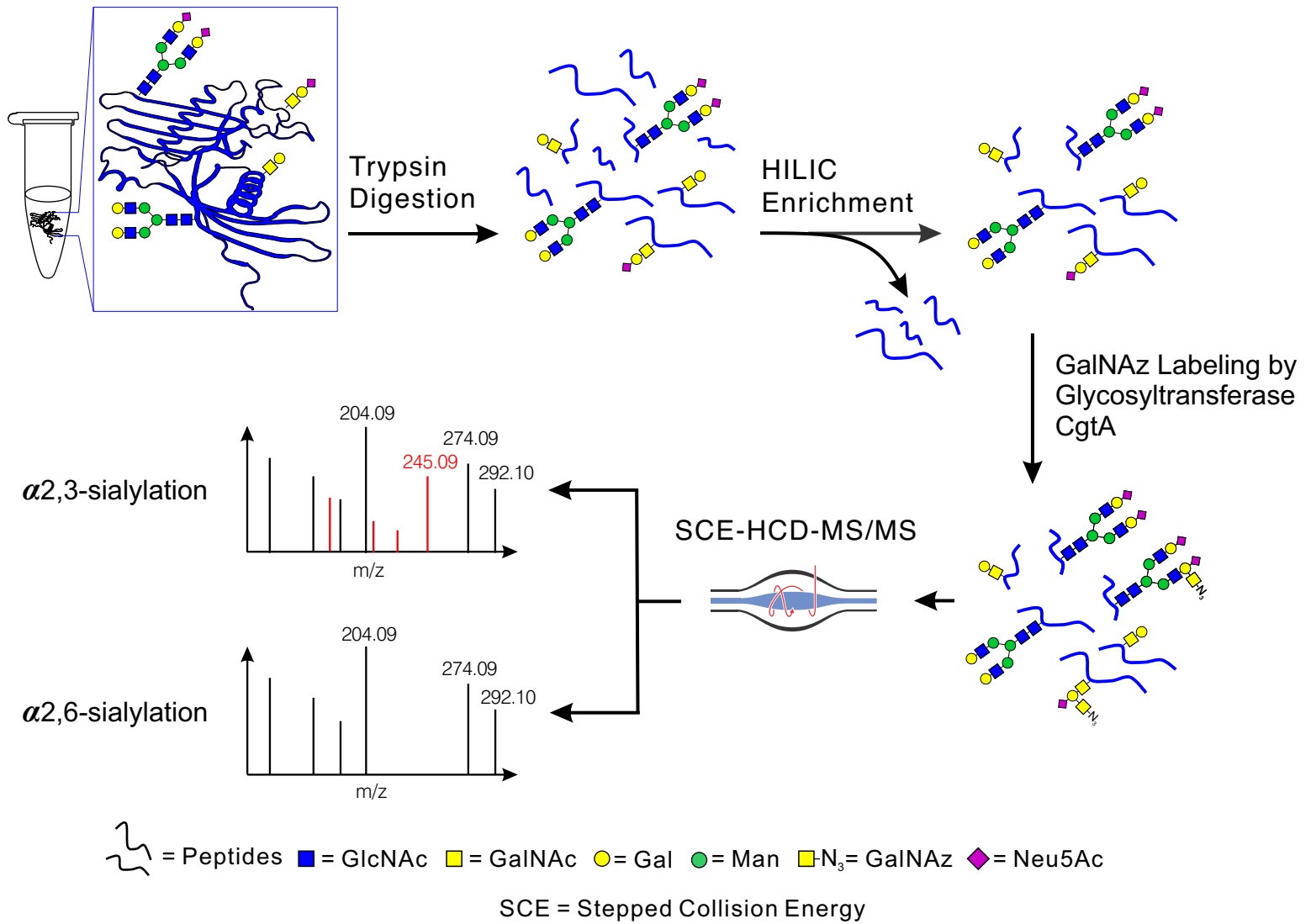
A



B



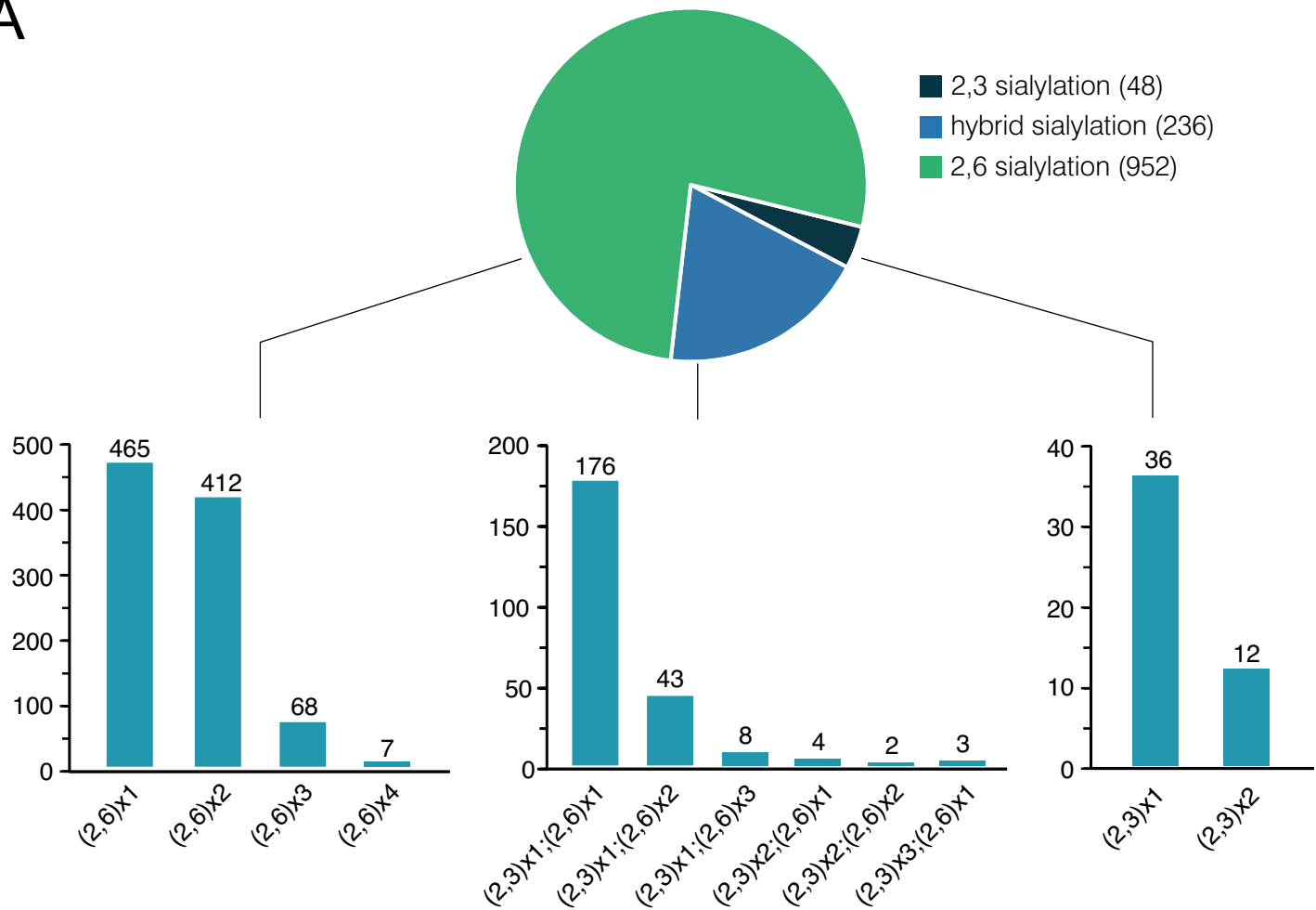
## Figure 4



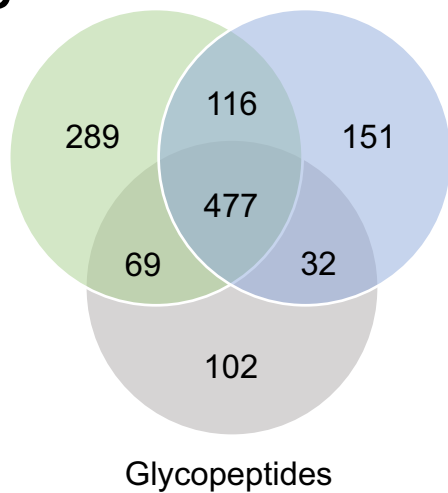


## Figure 5

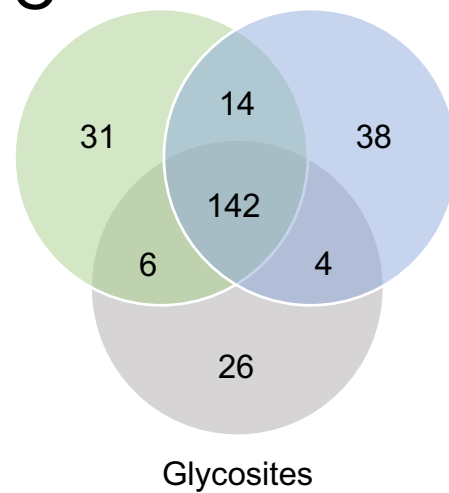
A



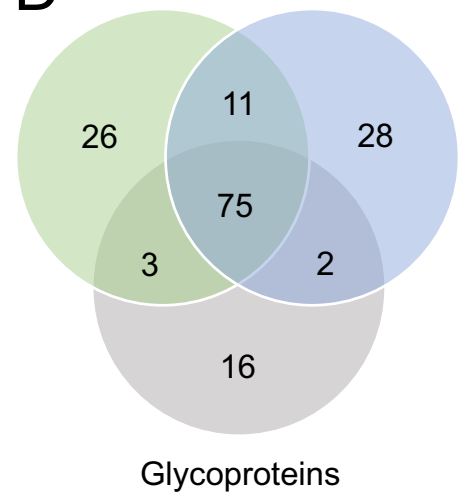
B



C



D



● Colon Cancer ● Liver Cancer ● Normal