Journal of proteome research

Article

Cite This: J. Proteome Res. 2019, 18, 3439−3446

pubs.acs.org/jpr

Highly Efficient Analysis of Glycoprotein Sialylation in Human Serum by Simultaneous Quantification of Glycosites and Site-Specific Glycoforms

Hongqiang Qin,[†,∥] Xuefang Dong,[†,∥] Jiawei Mao,[†,‡] Yao Chen,[†,‡] Mingming Dong,[†] Liming Wang,[§] Zhimou Guo,[*,†] Xinmiao Liang,[*,†] and Mingliang Ye[*,†]

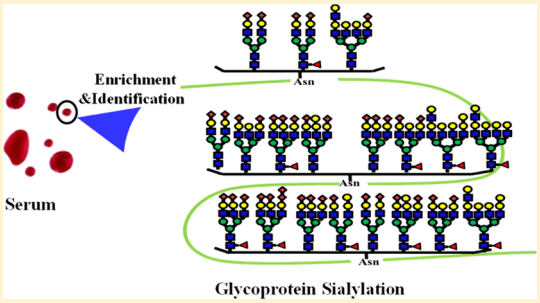
[†]CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, P. R. China

[‡]University of Chinese Academy of Sciences, Beijing 100049, China

[§]The Second Affiliated Hospital of Dalian Medical University, Dalian 116027, China

Supporting Information

ABSTRACT: Aberrant sialylation of glycoproteins is closely related to many malignant diseases, and analysis of sialylation has great potential to reveal the status of these diseases. However, in-depth analysis of sialylation is still challenging because of the high microheterogeneity of protein glycosylation, as well as the low abundance of sialylated glycopeptides (SGPs). Herein, an integrated strategy was fabricated for the detailed characterization of glycoprotein sialylation on the levels of glycosites and site-specific glycoforms by employing the SGP enrichment method. This strategy enabled the identification of up to 380 glycosites, as well as 414 intact glycopeptides corresponding to 383 site-specific glycoforms from only initial 6 μL serum samples, indicating the high sensitivity of the method for the detailed analysis of glycoprotein sialylation. This strategy was further employed to the differential analysis of glycoprotein sialylation between hepatocellular carcinoma patients and control samples, leading to the quantification of 344 glycosites and 405 site-specific glycoforms, simultaneously. Among these, 43 glycosites and 55 site-specific glycoforms were found to have significant change on the glycosite and site-specific glycoform levels, respectively. Interestingly, several glycoforms attached onto the same glycosite were found with different change tendencies. This strategy was demonstrated to be a powerful tool to reveal subtle differences of the macro- and microheterogeneity of glycoprotein sialylation.



KEYWORDS: *sialylation, human serum, glycosites, intact glycopeptides, site-specific glycoforms*

INTRODUCTION

Sialylation, as a specific form of glycosylation, plays important roles in the regulation of protein−protein interactions, and its aberrations are closely related to the immunological diseases, drug resistance, tumor invasiveness, and so forth.[1−5] Thus, it is of great interest to evaluate the changes of sialylation in cancer patients compared to those of healthy samples. It has been reported that the increased sialylation of glycan structures appeared to be indicative of cancer progression in the analysis of glycomics and glycosite occupancy.[6,7] As we all know, glycosylation could occur on multiple glycosites and many different sialylated glycan structures could be linked onto the same glycosite, which demonstrate the high macro- and microheterogeneity of sialylation.[8] Unfortunately, most studies focused on the analysis of glycan structures or glycosites,[7,9,10] which could only reveal the average results of glycans from different glycosites or glycosite occupancies from the total glycoforms, and the heterogeneous information of sialylation in detail was lost during the process.

Intact sialylated glycopeptides (SGPs) presented in the digest of serum samples or cell lysates contained the information of glycan structures, peptide sequences, and glycosite locations, which could allow us to determine site-specific sialylated glycoforms. Because SGPs coexist with huge amounts of high-abundance nonglycopeptides, the detection of SGPs from the complex protein digest is difficult, and they must be specifically enriched from the protein digest. Hydrazine chemistry has been widely used in the enrichment of N- and O-linked SGPs with high specificity by exploiting the selective oxidization of sialic acids.[11] However, the information of sialylation was lost during the oxidization processes; thus, it is not applicable to analyze the intact SGPs. Titanium dioxide (TiO_2) chromatography could enrich the SGPs without destroying the sialic acid structures.[12] Yet, this method suffers the disadvantages of limited specificity and efficiency. In our recent work, a hydrophilic smart material with switchable

Received: May 23, 2019
Published: August 5, 2019

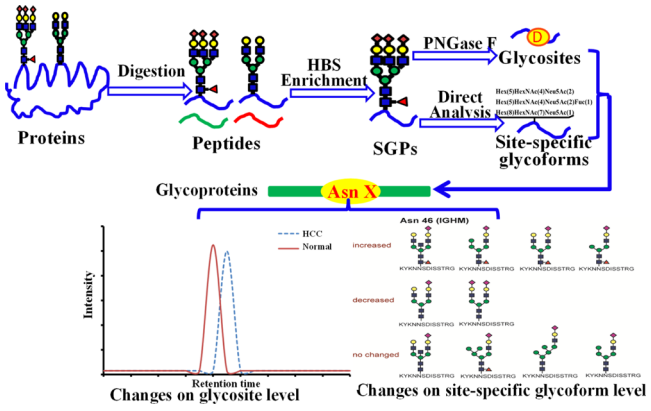
DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

surface charge (histidine-bonded silica, HBS) modified with histidine was reported to selectively enrich SGPs from complex biological samples.[13] This enrichment method was based on the high hydrophilicity and negative charge of SGPs, and the integrity of glycopeptides could be retained during the enriched process, which is fitted for the analysis of glycosites and site-specific glycoforms, simultaneously.

In addition to the SGP enrichment, it is also difficult to identify intact glycopeptides by mass spectrometry for the sialylation analysis on site-specific glycoform levels. In our lab, a method based on paired spectra was developed by integrating the spectra of both the deglycopeptides and intact glycopeptides.[10,14−16] The two types of spectra were matched by the delta retention time and the mass tolerance, and this method could enhance the identification efficiency of intact glycopeptides. Yet, the sialic acids could increase the retention time of intact glycopeptides in long gradient liquid chromatography tandem mass chromatography (LC−MS)/MS, which induced a high false-positive discovery rate (FDR) for the determination of intact SGPs. Up to now, other strategies, such as pGlyco2.0 and SugarQb, were also developed to identify intact glycopeptides.[17−19] Among them, pGlyco2.0 was reported to have good performance for the identification of intact glycopeptides, which does not rely on the retention of glycopeptides, and so the increased retention time of sialic acids has no influence on the FDR of intact SGPs. Thus, pGlyco2.0 is a good choice for the analysis of glycoprotein sialylation on site-specific glycoform levels.

As mentioned above, a detailed analysis of glycosites, as well as site-specific glycoforms, could provide potential biomarkers for the early diagnosis of human cancers, as well as enhance the specificity and sensitivity of the biomarkers. In this study, we presented an integrated workflow for the simultaneous analysis of glycosites and site-specific glycoforms by taking advantages of the SGP enrichment approach with high specificity and efficiency (shown in Scheme 1). This strategy enabled the detailed analysis of sialylation of glycoproteins in human serum samples on both levels, which depicted the precise macro- and microheterogeneity of protein sialylation. It was further applied to the differential analysis of glycoprotein sialylation between the hepatocellular carcinoma (HCC) and control serum samples. This method could be a powerful tool for the analysis of glycoprotein sialylation in human serum and other biological samples, which is expected to have great applications in the discovery of novel site-specific glycoform disease biomarkers.

Scheme 1. Workflow for the Comprehensive Analysis of Glycoprotein Sialylation on the Glycosite and Site-Specific Glycoform Levels



EXPERIMENTAL SECTION

Reagents and Materials

Formic acid (FA), dithiothreitol (DTT), iodoacetamide (IAA), PNGase F, and trypsin (TPCK-treated) were obtained from Sigma (St. Louis, MO). Acetonitrile (ACN), methanol, and ammonium hydroxide (25 wt % solution in water) of high-performance liquid chromatography grade were obtained from Merck (Darmstadt, Germany). Ammonium bicarbonate (NH_4HCO_3) and histidine were obtained from J&K Scientific (Beijing, China). (3-Glycidyloxypropyl) triethoxysilane (98%) was purchased from TCI (Tokyo, Japan). Human serum samples from 100 healthy volunteers were employed to test the identification efficiency of glycosites and site-specific glycoforms. For the relative quantification of glycosites and site-specific glycoforms, serum mixed from 10 healthy volunteers and 10 HCC patients were used as the control and HCC samples, respectively. All the human serum samples were stored at −80 °C before use. This study was approved by the Institutional Review Board of Dalian Medical University (Dalian, China). Informed consent was obtained from patients enrolled. Water used in all the experiments was purified by a Milli-Q water purification system (Millipore, Milford, MA, USA). Spherical silica (5 μm particle size, 100 Å pore size, and 300 m^2 g^{-1} surface area) was purchased from Fuji Silysia Chemical (Kasugai, Japan). C18 AQ beads (3 μm, 120 Å) for nano-LC−MS/MS was obtained from Michrom Bio Resources (Auburn, CA, USA).

Digestion of Human Serum Proteins

Human serum (10 μL) was mixed with 100 μL of 8 M urea in 50 mM ammonium bicarbonate (NH_4HCO_3, pH 8.2). The resultant solution was treated with 10 mM DTT for 2 h at 37 °C. Then, 20 μL of 100 mM IAA was added, and the mixture was incubated in the dark for 30 min at room temperature. The mixture was diluted to 1.0 mL with 50 mM NH_4HCO_3 and digested by trypsin at the enzyme-to-protein mass ratio of 1:25 and incubated at 37 °C overnight. To remove the salt and matrix, the tryptic mixtures were loaded onto a homemade C18 tip and eluted with 75% ACN/0.1% FA solution for further process.

Enrichment of Glycopeptides from Protein Digests

The human serum tryptic digest (5 μL of the initial human serum) was first mixed with 5.0 mg of material (HBS materials prepared in the report[13]) in 0.50 mL of 75% ACN/0.1% FA and shaken for 1 h followed by centrifugation at 10 000 *g* for 2 min. After this, the supernatant was removed and the precipitate was washed with 75% ACN/0.1% FA (0.50 mL), 70% ACN/0.1% FA (0.50 mL), and 70% ACN/0.05% HAc (0.50 mL) successively and centrifuged at 10 000*g* for 2 min. Then, the enriched SGPs were eluted with 0.50 mL of 40% ACN/5 mM NH_4HCO_3 (pH 8.2). After centrifugation at 10 000*g* for 2 min, the supernatant was collected and dried for the analysis of glycosites and site-specific glycoforms by nano-LC−MS/MS. For the analysis of glycosites, the SGPs were treated by using PNGase F to remove the glycans in 20 mM NH_4HCO_3 (pH 8.2) at 37 °C overnight. The solutions were dried and dissolved in 0.1% FA for MS analysis.

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

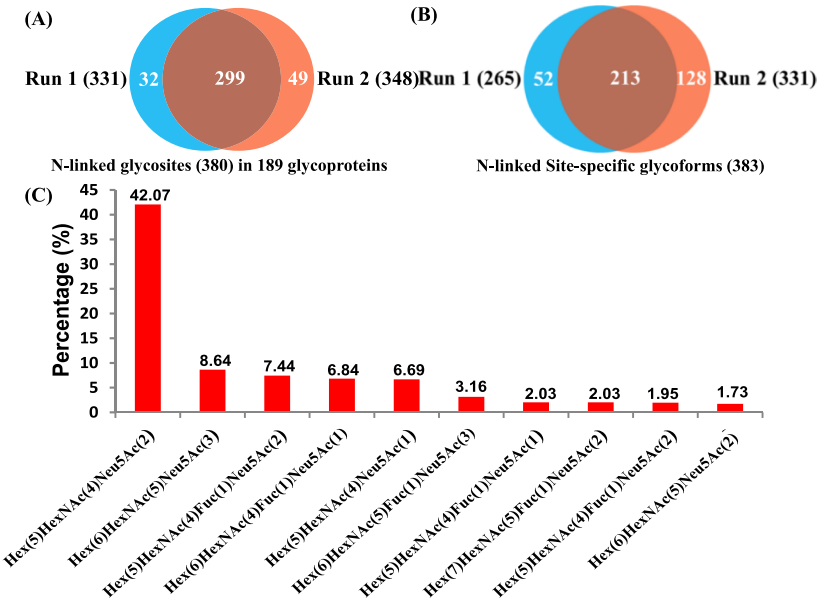


Figure 1. Identification of sialylated N-linked glycosites (A) and site-specific glycoforms (B) from normal human serum samples pooled with 100 individuals. (C) Distributions of the top 10 high-abundance glycoforms identified in the intact glycopeptides.

Mass Spectrometric Analysis of Glycosites and Site-Specific Glycoforms

The LC−MS/MS analyses were performed on a Q-Exactive mass spectrometer (Thermo, San Jose, CA), equipped with an Ultimate 3000 system (Thermo, San Jose, CA) for separation. The LC−MS/MS system contained a 4 cm C18 capillary trap column (200 μm i.d., C18 AQ beads (5 μm, 120 Å)) and a 12 cm C18 capillary analysis column (75 μm i.d., C18 AQ beads (3 μm, 120 Å)). The deglycopeptides and intact glycopeptides obtained from 1.5 μL of original human serum were loaded for one reverse-phase liquid chromatography (RPLC)−MS/MS run, respectively. The sample was loaded for 10 min at the flow rate of 4.0 μL/min using 0.1% FA in water. The gradient was generated as follows: from 3 to 6% buffer B for 5 min, from 6 to 35% buffer B for 165 min, from 35 to 45% buffer B for 25 min, from 45 to 90% buffer B for 15 min, and held on 90% buffer B for 10 min. The separation system was equilibrated by buffer A for 10 min. The mobile phases A and B were 2% ACN and 0.1% trifluoroacetic acid (TFA) and 98% ACN and 0.1% TFA, respectively. The temperature of the ion-transfer capillary was 275 °C, and the normalized collision energy was set to 27 ± 3% for the analysis of glycosites and 25 ± 5% for the analysis of intact glycopeptides [higher energy collisional dissociation (HCD) mode]. The mass resolution was set to 70 000 for full MS and 17 500 for HCD MS/MS. Full survey scan MS was acquired from m/z 400 to 2000, and 15 most intense ions were selected for MS/MS scan. The dynamic exclusion was set as follows: repeat count, 1; duration, 30 s; exclusion list size, 500; exclusion duration, 30 s. All the samples, including intact glycopeptides and deglycopeptides were analyzed for three runs.

Data Analysis

The mass spectrometric data acquired by Q-Exactive MS were converted into a peaklist file (.mgf and .MZxml) using Protein Discovery and MSconvert. For the identification and quantification of glycosites, the raw data were searched by using label-free quantification (LFQ) in Maxquant (1.6.1.0).[20] The database was obtained from UniProtKB of human database with 20 193 proteins reviewed. The parameters used for searching were as follows: mass tolerances were 10 and 50 ppm for the precursors and fragments, respectively; for trypsin-digested samples, enzyme specificity was set to KR/P with up to two missed cleavages; cysteine residue was set as a static modification of +57.0215 Da; and Asn deamidation (+0.9840 Da) and methionine oxidation (+15.9949 Da) were set as variable modifications. "Match between runs" was employed to the quantitative analysis of glycosites between the two sample groups, and the retention time window was set as 3 min. The obtained quantitative data were further analyzed by using Perseus software (version 1.5.8.5).[21] For the quantification of glycosites, p-value < 0.05 and s_0 = 0.1 were adopted as the cut-off values of the glycosites with significant change. Because of the missing value, the glycosites quantified without p-value were removed in this work. In addition, the only observed glycosites in the HCC group and control group were considered as "upregulated" and "downregulated" in HCC, respectively.

For the determination of site-specific glycoforms, the raw data of intact glycopeptides were directly searched by using pGlyco2.0.[22] The search parameters were used as follows: mass tolerances were 10 and 50 ppm for the precursors and fragments, respectively; for trypsin-digested samples, the enzyme specificity was set to KR/P with up to three missed cleavages; cysteine residue was set as a static modification of +57.0215 Da; methionine oxidation (+15.9949 Da) was set as a variable modification. After the database search, the total FDR was set as <1%, and the spectra were collected for the identification of intact glycpeptides and site-specific glycoforms. The peptides with the same backbones and glycan compositions were combined as the same intact glycopeptides, including some glycopeptides with the modifications of methionine oxidation, and so forth.

For the quantification of site-specific glycoforms, the intensity of intact glycopeptides was extracted from six raw data by using pQuant using the parameter list in the Supporting Information. And cut-off of the similarity of the isotope was selected as >0.90 as previous report, and the

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

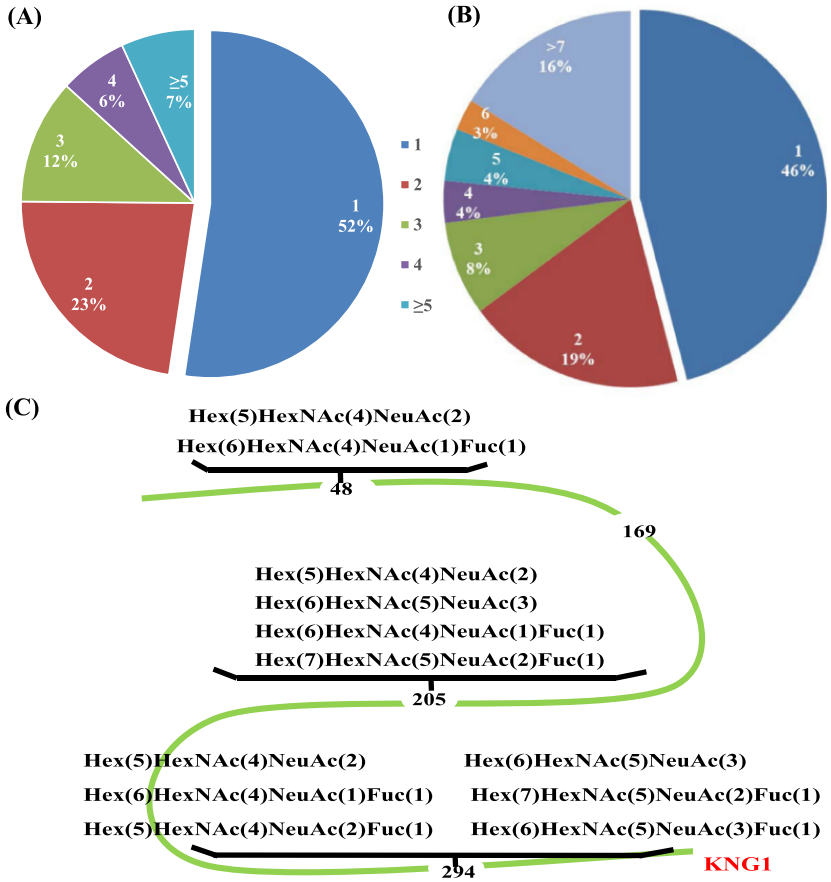


Figure 2. Distributions of glycoproteins with different numbers of glycosites (A) and glycosites with different numbers of site-specific glycoforms (B). (C) Example for the glycosites and glycoforms of KNG1 identified from human serum sample by using the integrated platform.

intensity was labeled as missing values for the intact glycopeptides with similarity lower than 0.90.[23] Then, the intact glycopeptides with the highest intensity were employed to the quantitative analysis of the site-specific glycoforms, and the ratios between the samples were utilized to screen the changed site-specific glycoforms. For the site-specific glycoforms, the p-value < 0.05 and ratio (>2 or <0.5) were adopted as the cut-off values of the site-specific glycoforms with significant change. The solely extracted site-specfic glycoforms in the HCC or control groups were cosidered as "upregulated" and "downregulated", respectively. Because of the missing value, the site-specific glycoforms quantified without p-value were removed in this work. The raw data have been loaded in the iProx, named as PXD009654 (https://www.iprox.org//page/project.html?id=IPX0001208000).

RESULTS AND DISCUSSION

The performance of this integrated workflow (Scheme 1) for the analysis of glycosites and site-specific sialylated glycoforms was investigated by analyzing the human serum sample pooled from 100 healthy individuals. The human serum digest was subjected to HBS enrichment, which could selectively enrich the SGPs and remove the high-abundance nonglycopeptides, as well as non-SGPs. This is the critical step for the in-depth analysis of glycoprotein sialylation. Then, the enriched SGPs were divided into two aliquots: one aliquot was treated with PNGase F for deglycosylation (refer to deglycopeptides) and the other one was kept as intact glycopeptides. The above two samples were subjected to LC−MS/MS analysis, respectively. The analysis of the deglycopeptides by two LC−MS/MS replicate runs resulted in the identification of 380 N-linked glycosites corresponding to 189 glycoproteins (Figure 1A, Table S1). For this analysis, only 3.0 μL of human serum samples (1.5 μL serum/run) were consumed. Compared with the previous method of SGP enrichment by using Ti-IMAC beads, twofold more glycosites were identified.[24] Compared with the hydrophilic interaction LC (HILIC) method, 72% more glycosites were identified from the same samples. Additionally, a high specificity of glycopeptide enrichment was obtained from human serum samples by using the HBS enrichment as evidenced by the fact that more than 65% of peptides were identified as glycopeptides in this study.[25] The improved enrichment specificity could significantly decrease the false-positive identification caused by the hydrolysis of Gln and Asn.[26] As shown in Figure S1, over 95% (361/380) of the glycosites identified by this method have been reported in the UniProt database, indicating the high confidence of the identified glycosites from human serum samples.

For further determination of the intact SGPs, the MS/MS spectra of intact glycopeptides were interpreted by using the pGlyco2.0 software.[16] As shown in Figure 1B, a total number of 405 intact glycopeptides and 383 site-specific glycoforms were identified in two replicate runs (1.5 μL of original human serum sample per run) (Table S2). It was found that nearly 99% (378/383) of the site-specific glycoforms contained terminal sialic acids, and all of the top 10 high-abundance glycans were sialylated (Figure 1C). Interestingly, a similar percentage of fucosylated glycopeptides was obtained by using

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

HBS and HILIC enrichments, whereas only 66.30% of the glycopeptides were determined as SGPs enriched by using the HILIC method (Figure S2). The above results demonstrated that the HBS method could specifically enrich SGPs from the peptide mixture.

The feature of protein glycosylation is its high macro-heterogeneity (one glycoprotein could have several glycosites) and high microheterogeneity (each glycosite could be occupied by multiple types of glycan structures). Indeed, we found that around 50% (90/189) of the identified proteins had more than two glycosites (Figure 2A) and 54% of the glycosites were occupied by more than two glycoforms, including 16% of the glycosites with even more than seven glycoforms (Figure 2B). It was also found that some proteins were hyperglycosylated. For example, there were five glycosites identified in hemopexin (HEMO), and as many as 23 site-specific glycoforms were detected on two of the glycosites (Asn187 and Asn453). Multiple glycosites and diverse glycoforms make the protein glycosylation extremely heterogeneous. Kininogen-1 (KNG1) is one of the inhibitors of thiol proteases, which plays important roles in blood coagulation. Only four site-specific glycoforms on two glycosites (Asn48 and Asn169) were identified for this protein in a recent study.[27] However, in our results, all of the four glycosites (Asn48, Asn169, Asn205, and Asn294) in the UniProt database were covered (Figure 2C), and as many as 12 site-specific glycoforms were identified on three of the glycosites (Asn48, Asn205, and Asn294), which were all sialylated as HEMO, indicating the high specificity of the SGP enrichment. Compared with the previous work, many more glycosites could be obtained from less than 1/20 of the amount serum samples (40 μL of human plasma in previous results).[28] Take the glycosylation on haptoglobin as another example. There were four N-linked glycosites (Asn184, Asn207, Asn211, and Asn241) reported in the UniPort database, and all the four glycosites were covered in our data, indicating the high coverage of glycosite analysis by using this method. Compared with the previous work based on the lectin and SAX−ERLIC enrichment method, more than 2 times of intact glycopeptides were identified from 1/5 of the amount of human samples.[29] The above results indicated that the strategy presented in this study could identify the glycosites and site-specific glycoforms with high sensitivity, which effectively revealed the macro- and microheterogeneity of glycoprotein sialylation.

This strategy was further applied to the differential analysis of the HCC and control samples. For the relative quantification of glycosites and site-specific glycoforms, serum samples mixed from 10 healthy volunteers and 10 HCC patients were used as the control and HCC samples, respectively. The enriched deglycopeptides from the two samples were run three replicates by using one-dimensional RPLC−MS/MS. Totally, 389 and 434 glycosites were identified from the HCC and normal samples, respectively (Figure 3). Bringing together, 344 glycosites corresponding to 188 glycoproteins were identified from these two samples (Table S3A). It obtained 43 glycosites in 39 glycoproteins significantly changed on sialylated glycosite level between HCC and control human serum (Tables 1 and S3B). Among them, 18 glycosites were significantly increased, whereas 25 glycosites were notably decreased in the HCC serum sample. Immunoglobulins play important roles in individual immunity to pathogens and HCC.[2] In comparison with the normal human serum sample, five glycosites in immunoglobulins were

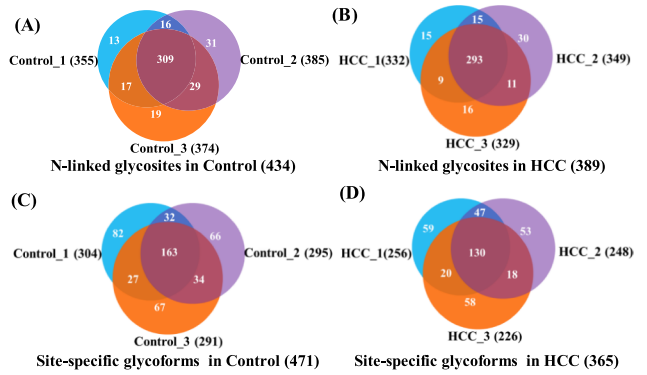


Figure 3. N-linked glycosites identified from control (A) and HCC human serum (B) in three runs; overlap of sialylated site-specific glycoforms determined from control (C) and HCC human serum (D) in three runs.

Table 1. Number of Sialylated N-Glycosites Significantly Changed between HCC and Control-Pooled Human Serum Samples

	glycosites	corresponding glycoproteins
upregulation in HCC serum[a]	18	16
downregulation in HCC serum[b]	25	24
total	43	39(1[c])

[a]"Upregulation in HCC" includes glycosites with p-value < 0.05 and difference > 0.5 and those exclusively observed in the HCC serum in at least three runs. [b]"Downregulation in HCC" includes glycosites with p-value < 0.05 and difference < 0.5 and those exclusively observed in the control serum in at least three runs. [c]One protein has both up- and downregulated sites.

observed, with a significant increase in HCC. One site, Asn177 on Ig γ-chain C region 4 (IGHG4), even increased as many as 16-fold in HCC serum samples. Many significantly changed glycoproteins have been reported to be closely related with HCC, such as Galectin 3 binding protein (Gal-3BP),[30] KNG1,[31] and so forth. Interestingly, there was a protein (CLU) that contained both one increased glycosite (Asn86) and one decreased glycosite (Asn291) in the HCC serum. As these two sites presented on the same protein, their difference must be because of the different sialylation occupation for different disease states.

The glycosite quantification using deglycosylated peptide analysis revealed the total site occupancies of all the glycoforms but not the occupancies of site-specific glycoforms. To further quantify the site-specific glycoforms, the enriched intact SGPs from the two samples were also analyzed by LC−MS/MS. pGlyco2.0 was employed to determine the intact SGPs enriched from the two pooled samples, yielding the identification of 471 and 365 site-specific glycoforms from only one drop (1.5 μL/run, total 4.5 μL/sample) of HCC and control human serum (Figure 3).

The distributions of the identified intact glycopeptides containing sialic acids and fucosylated glycan structures were also investigated. As shown in Figure S3, more than 99% of the site-specific glycoforms were determined as sialylation glycopeptides in both the samples. However, no obvious difference was observed for the ratios of sialylation and fucosylation between these two samples. Additionally, the spectra count distributions of glycosites bearing the same glycan structures were also investigated, and only a few

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

differences were observed between the two samples (Figure S4 and Table S4). It also should be mentioned that the glycan abundances in glycomics were the summed values of the same glycans onto all glycosites of glycoproteins. Thus, their abundance changes depend heavily on the changes of glycans attached onto high-abundance proteins, which could mask the detailed changes of glycoforms attached onto the proteins with low abundance. To address this issue, the site-specific glycoforms were further quantified to reveal the difference in the microheterogeneity of glycosylation between the two samples.

The quantitative analysis of site-specific glycoforms was achieved by the quantification of intact SGPs. The LFQ method was employed to quantify the site-specific glycoforms by using pQuant, and the intact SGPs with the highest intensity were adopted to evaluate the change tendency of the site-specific glycoforms. There were a total number of 560 site-specific glycoforms identified, and the intensity of the intact glycopeptides was obtained by using pQuant.[23] First, the repeatability of the quantitative analysis of site-specific glycoforms was evaluated, and a correlation coefficient of 0.85 was obtained for the triple runs from the control sample group, indicating the good repeatability for quantification between the different groups (Figure S5). Totally, 405 site-specific glycoforms were quantified between the two groups of samples (Table S5). Further, 19 sialylated site-specific glycoforms attached onto 14 glycosites were significantly increased, whereas 36 sialylated site-specific glycoforms onto 33 glycosites were obviously decreased in the HCC samples (Tables 2 and S5B). Among them, there were several glycosites

Table 2. Number of Sialylated Site-Specific Glycoforms Significantly Changed between HCC and Control-Pooled Human Serum Samples

	site-specific glycoforms	corresponding glycosites
upregulation in HCC serum[a]	19	14
downregulation in HCC serum[b]	36	33
total	55	41(6[c])

[a]"Upregulation in HCC" includes the site-specific glycoforms with p-value < 0.05 (ratio > 2) and those exclusively observed in the HCC serum in more than two runs. [b]"Downregulation in HCC" contains the site-specific glycoforms with p-value < 0.05 (ratio < 0.5) and those solely observed in the normal human serum in the two runs. [c]Six glycosites have both up- and downregulated site-specific glycoforms.

with both up- and downregulated site-specific glycoforms. For example, the glycoform of Hex(5)HexNAc(4)NeuAc(1)-Fuc(3) attached onto Asn126 in HPTR was upregulated in HCC samples, whereas the glycoform of Hex(7)HexNAc(5)-NeuAc(2)Fuc(2) attached onto the same glycosite was downregulated in the HCC sample (Figure 4A). There were also 10 glycoforms without significant changes between the two groups of samples. The site-specific glycoforms attached onto immune glycoproteins were also found to be significantly changed. For example, the glycosites and glycoforms of IgG (Asn144) were found to be significantly up-expressed from the HCC serum samples. As previously reported, the glycosylation of IgG has been used for the diagnosis of several cancers, such as breast cancer,[32] ovarian cancer,[33] and fucosylation of IgG for the HCC.[34] Thus, the glycosylation of IgG could be potential biomarkers for malignant tumors. Additionally, the same

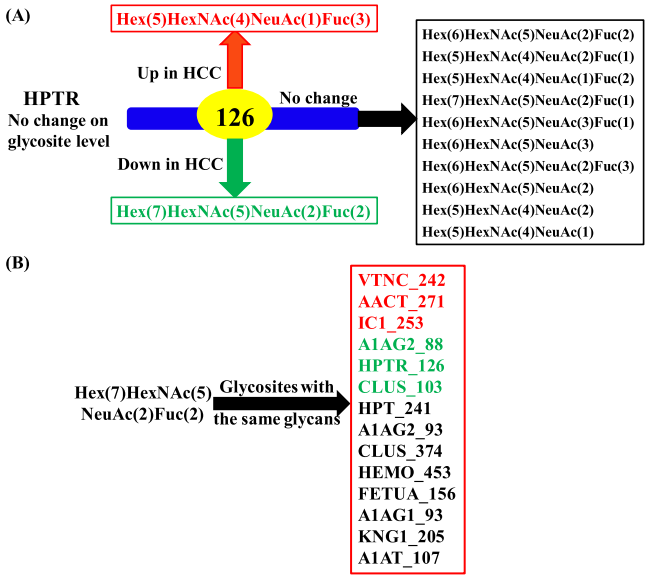


Figure 4. (A) Different change tendencies of the glycans attached onto Asn126 in HPTR (HCC vs Control). (B) Change tendencies of the same glycoform of Hex(7)HexNAc(5)Neu5Ac(2)Fuc(2) on different sites of various proteins (red indicates upregulation in HCC; green indicates downregulation in HCC; black indicates no significant change; the numbers indicate the sites).

glycans attached onto different proteins may play quite distinct roles, and the glycans could show different tendencies between the proteins. For example, Hex(7)HexNAc(5)Neu5Ac(2)-Fuc(2) was one of the glycans attached onto Asn126 in HPTR with significant upregulation. This sialylated glycoform was also present on other 11 glycosites of other proteins. These glycoforms have different change tendencies, including three upregulated and three downregulated glycoforms, as well as six sialylated site-specific glycoforms without significant change (Figure 4B). These above results indicated the high micro-heterogeneity of protein sialylation on site-specific glycoform levels. Thus, a comparison was made between the sialylation on glycosite and site-specific glycoform levels. For the molecular function gene ontology (GO) terms enriched for the glycoproteins, there was no significant change between the glycoproteins from the total glycosites, significant change on the glycosite level, and significant change on the site-specific glycoform level (Figure S6). It needs a detailed analysis of glycosite occupancy and site-specific glycoforms. In addition, it was found that similar change tendency was obtained for several glycosites and site-specific glycoforms of sialylation (shown in Table S6). However, there were still many glycoproteins with different change tendencies on the two levels, caused by the heterogeneity of protein sialylation. Clearly, the quantification on glycosites and site-specific glycoforms could more accurately reveal the detailed regular function of glycoprotein sialylation. In this work, the integrated platform has been fabricated for the analysis of protein sialylation, and the sialylated glycosites and site-specific glycoforms were simultaneously quantified from the pooled HCC and control serum samples. By using this method, several site glycoforms were obtained with significant change. These site-specific glycoforms should be further confirmed by using a cohort of individual samples for diagnostic purposes. In the future work, the analysis of site-specific glycoforms with high

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

throughput will be developed for the confirmation of the datasets by using individual samples.

Combining all the data sets of the pooled serum samples including the initial optimization analysis, a total number of 714 site-specific glycoforms corresponding to 759 intact glycopeptides, including only about 0.5% (4/759) of the non-SGPs, were identified from human serum (Table S7A,B). This is, to the best of our knowledge, one of the largest datasets of sialylated site-specific glycoforms from human serum samples, which could be used for a detailed study of glycoprotein sialylation in human serum samples.

CONCLUSIONS

In summary, we presented an integrated strategy for the analysis of sialylation on the glycosite and site-specific glycoform levels, which could map the macro- and microheterogeneity of glycoprotein sialylation, simultaneously. A total number of 714 site-specific glycoforms corresponding to 759 intact glycopeptides were identified from human serum by using the integrated strategy. Several glycosites were observed with different change tendencies between relative site-occupancy and site-specific glycoform levels. This platform has been proved as a powerful tool for the in-depth analysis of macro- and microheterogeneity of glycosylation in human serum. Thus, it has great potential to screen novel biomarkers for the early diagnosis of human cancers.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.9b00332.

Sialylated glycosites and intact glycopeptides identified from the human serum trypsin digest; quantification results of glycosites and site-specific glycoforms between HCC and control human serum; list of glycan structures identified from intact glycopeptides; total intact glycopeptides and site-specific glycoforms identified from human serum; overlap of glycosites identified between our data and the UniProt database; distributions of site-specific glycoforms containing sialic acids and fucosylated glycan structures; numbers of unique intact glycopeptides bearing the same glycans; repeatability of the quantitative analysis of site-specific glycoforms; and molecular function GO terms enriched for the glycoproteins obtained on different levels (PDF)

Glycosites enriched by using HBS method from human serum tryptic digest (XLSX)

Intact glycopeptides enriched by using HBS method from human serum tryptic digest (XLSX)

Quantification of glycosites between HCC and normal human serum samples (XLSX)

Count of glycan structures identifed between healthy and HCC serum samples (XLSX)

Site-specific glycoforms quantified between HCC and control serum samples by using pQuant (XLSX)

Comparsion of significant changes between glycosites and site-specific glycoforms (XLSX)

Total number of site-specific glycoforms identified from human serum samples (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: guozhimou@dicp.ac.cn. Phone: +86-411-84379539 (Z.G.).
*E-mail: lianagxm@dicp.ac.cn. Phone: +86-411-84379519 (X.L.).
*E-mail: mingliang@dicp.ac.cn. Phone: +86-411-84379610 (M.Y.).

ORCID

Hongqiang Qin: 0000-0002-7508-0872
Xinmiao Liang: 0000-0001-5802-1961
Mingliang Ye: 0000-0002-5872-9326

Author Contributions

‖H.Q. and X.D. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported, in part, by funds from the China State Key Basic Research Program Grants (2018YFC0910302, 2016YFA0501402), the National Natural Science Foundation of China (21535008, 81430072, 81361128, and 21775146), the LiaoNing Revitalization Talents Program from the Liaoning province of China, and the innovation program of Dalian Institute of Chemical Physics, CAS (DICP TMSR201601, DICP&QIBEBT UN201802), and M.Y. is a recipient of the National Science Fund of China for Distinguished Young Scholars (21525524).

DEDICATION

Dedicated to the 70th anniversary of Dalian Institute of Chemical Physics, CAS.

REFERENCES

(1) Tsuboi, S.; Hatakeyama, S.; Ohyama, C.; Fukuda, M. Two opposing roles of O-glycans in tumor metastasis. *Trends Mol. Med.* **2012**, *18*, 224−232.

(2) Bhide, G. P.; Colley, K. J. Sialylation of N-glycans: mechanism, cellular compartmentalization and function. *Histochem. Cell Biol.* **2017**, *147*, 149−174.

(3) Harre, U.; Lang, S. C.; Pfeifle, R.; Rombouts, Y.; Frühbeißer, S.; Amara, K.; Bang, H.; Lux, A.; Koeleman, C. A.; Baum, W.; Dietel, K.; Gröhn, F.; Malmström, V.; Klareskog, L.; Krönke, G.; Kocijan, R.; Nimmerjahn, F.; Toes, R. E. M.; Herrmann, M.; Scherer, H. U.; Schett, G. Glycosylation of immunoglobulin G determines osteoclast differentiation and bone loss. *Nat. Commun.* **2015**, *6*, 6651.

(4) Debruyne, E. N.; Delanghe, J. R. Diagnosing and monitoring hepatocellular carcinoma with alpha-fetoprotein: New aspects and applications. *Clin. Chim. Acta* **2008**, *395*, 19−26.

(5) Block, T. M.; Comunale, M. A.; Lowman, M.; Steel, L. F.; Romano, P. R.; Fimmel, C.; Tennant, B. C.; London, W. T.; Evans, A. A.; Blumberg, B. S.; Dwek, R. A.; Mattu, T. S.; Mehta, A. S. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 779−784.

(6) Kyselova, Z.; Mechref, Y.; Kang, P.; Goetz, J. A.; Dobrolecki, L. E.; Sledge, G. W.; Schnaper, L.; Hickey, R. J.; Malkas, L. H.; Novotny, M. V. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin. Chem.* **2008**, *54*, 1166−1175.

(7) Zhang, Z.; Sun, Z.; Zhu, J.; Liu, J.; Huang, G.; Ye, M.; Zou, H. High-Throughput Determination of the Site-Specific N-Sialoglycan Occupancy Rates by Differential Oxidation of Glycoproteins

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446

Followed with Quantitative Glycoproteomics Analysis. *Anal. Chem.* **2014**, *86*, 9830−9837.
(8) Dell, A.; Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* **2001**, *291*, 2351−2356.
(9) Yabu, M.; Korekane, H.; Miyamoto, Y. Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology* **2014**, *24*, 542−553.
(10) Chen, R.; Tan, Y.; Wang, M.; Wang, F.; Yao, Z.; Dong, L.; Ye, M.; Wang, H.; Zou, H. Development of Glycoprotein Capture-Based Label-Free Method for the High-throughput Screening of Differential Glycoproteins in Hepatocellular Carcinoma. *Mol. Cell. Proteomics* **2011**, *10*, M110.006445.
(11) Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **2009**, *6*, 809−811.
(12) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5*, 1974−1982.
(13) Dong, X.; Qin, H.; Mao, J.; Yu, D.; Li, X.; Shen, A.; Yan, J.; Yu, L.; Guo, Z.; Ye, M.; Zou, H.; Liang, X. In-Depth Analysis of Glycoprotein Sialylation in Serum Using a Dual-Functional Material with Superior Hydrophilicity and Switchable Surface Charge. *Anal. Chem.* **2017**, *89*, 3966−3972.
(14) Wu, J.; Qin, H.; Li, T.; Cheng, K.; Dong, J.; Tian, M.; Chai, N.; Guo, H.; Li, J.; You, X.; Dong, M.; Ye, M.; Nie, Y.; Zou, H.; Fan, D. Characterization of site-specific glycosylation of secreted proteins associated with multi-drug resistance of gastric cancer. *Oncotarget* **2016**, *7*, 25315−25327.
(15) Chen, R.; Seebun, D.; Ye, M.; Zou, H.; Figeys, D. Site-specific characterization of cell membrane N-glycosylation with integrated hydrophilic interaction chromatography solid phase extraction and LC-MS/MS. *J. Proteomics* **2014**, *103*, 194−203.
(16) Cheng, K.; Chen, R.; Seebun, D.; Ye, M.; Figeys, D.; Zou, H. Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J. Proteomics* **2014**, *110*, 145−154.
(17) Zhou, J.; Yang, W.; Hu, Y.; Höti, N.; Liu, Y.; Shah, P.; Sun, S.; Clark, D.; Thomas, S.; Zhang, H. Site-Specific Fucosylation Analysis Identifying Glycoproteins Associated with Aggressive Prostate Cancer Cell Lines Using Tandem Affinity Enrichments of Intact Glycopeptides Followed by Mass Spectrometry. *Anal. Chem.* **2017**, *89*, 7623−7630.
(18) Li, Q. K.; Chen, L.; Ao, M.-H.; Chiu, J. H.; Zhang, Z.; Zhang, H.; Chan, D. W. Serum Fucosylated Prostate-specific Antigen (PSA) Improves the Differentiation of Aggressive from Non-aggressive Prostate Cancers. *Theranostics* **2015**, *5*, 267−276.
(19) Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Dürnberger, 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*, 538−542.
(20) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513−2526.
(21) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731−740.
(22) 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*, 438.
(23) Liu, C.; Song, C.-Q.; Yuan, Z.-F.; Fu, Y.; Chi, H.; Wang, L.-H.; Fan, S.-B.; Zhang, K.; Zeng, W.-F.; He, S.-M.; Dong, M.-Q.; Sun, R.-X. pQuant Improves Quantitation by Keeping out Interfering Signals and Evaluating the Accuracy of Calculated Ratios. *Anal. Chem.* **2014**, *86*, 5286−5294.
(24) Zhu, J.; Wang, F.; Cheng, K.; Dong, J.; Sun, D.; Chen, R.; Wang, L.; Ye, M.; Zou, H. A simple integrated system for rapid analysis of sialic-acid-containing N-glycopeptides from human serum. *Proteomics* **2013**, *13*, 1306−1313.
(25) Zhu, J.; Wang, F.; Chen, R.; Cheng, K.; Xu, B.; Guo, Z.; Liang, X.; Ye, M.; Zou, H. Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. *Anal. Chem.* **2012**, *84*, 5146−5153.
(26) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K. Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation. *Mol. Cell. Proteomics* **2011**, *10*, O111.009381.
(27) Choi, N. Y.; Hwang, H.; Ji, E. S.; Park, G. W.; Lee, J. Y.; Lee, H. K.; Kim, J. Y.; Yoo, J. S. Direct analysis of site-specific N-glycopeptides of serological proteins in dried blood spot samples. *Anal. Bioanal. Chem.* **2017**, *409*, 4971−4981.
(28) Zhou, H.; Froehlich, J. W.; Briscoe, A. C.; Lee, R. S. The GlycoFilter: a simple and comprehensive sample preparation platform for proteomics, N-glycomics and glycosylation site assignment. *Mol. Cell. Proteomics* **2013**, *12*, 2981−2991.
(29) Kontro, H.; Joenväärä, S.; Haglund, C.; Renkonen, R. Comparison of sialylated N-glycopeptide levels in serum of pancreatic cancer patients, acute pancreatitis patients, and healthy controls. *Proteomics* **2014**, *14*, 1713−1723.
(30) Liu, T.; Liu, D.; Liu, R.; Jiang, H.; Yan, G.; Li, W.; Sun, L.; Zhang, S.; Liu, Y.; Guo, K. Discovering potential serological biomarker for chronic Hepatitis B Virus-related hepatocellular carcinoma in Chinese population by MAL-associated serum glycoproteomics analysis. *Sci. Rep.* **2017**, *7*, 38918.
(31) Wang, M.; Sanda, M.; Comunale, M. A.; Herrera, H.; Swindell, C.; Kono, Y.; Singal, A. G.; Marrero, J.; Block, T.; Goldman, R.; Mehta, A. Changes in the Glycosylation of Kininogen and the Development of a Kininogen-Based Algorithm for the Early Detection of HCC. *Cancer Epidemiol., Biomarkers Prev.* **2017**, *26*, 795−803.
(32) Gebrehiwot, A. G.; Melka, D. S.; Kassaye, Y. M.; Gemechu, T.; Lako, W.; Hinou, H.; Nishimura, S.-I. Exploring serum and immunoglobulin G N-glycome as diagnostic biomarkers for early detection of breast cancer in Ethiopian women. *BMC Cancer* **2019**, *19*, 588.
(33) Saldova, R.; Royle, L.; Radcliffe, C. M.; Abd Hamid, U. M.; Evans, R.; Arnold, J. N.; Banks, R. E.; Hutson, R.; Harvey, D. J.; Antrobus, R.; Petrescu, S. M.; Dwek, R. A.; Rudd, P. M. Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **2007**, *17*, 1344−1356.
(34) Mehta, A.; Block, T. M. Fucosylated glycoproteins as markers of liver disease. *Dis. Markers* **2008**, *25*, 259−265.

DOI: 10.1021/acs.jproteome.9b00332
J. Proteome Res. 2019, 18, 3439−3446