Journal of proteome research

Article

pubs.acs.org/jpr

LC−MS Profiling of N-Glycans Derived from Human Serum Samples for Biomarker Discovery in Hepatocellular Carcinoma

Tsung-Heng Tsai,[†] Minkun Wang,[†,¶] Cristina Di Poto,[†,¶] Yunli Hu,[‡,¶] Shiyue Zhou,[‡,¶] Yi Zhao,[†,¶] Rency S. Varghese,[†] Yue Luo,[†] Mahlet G. Tadesse,[§] Dina Hazem Ziada,[∥] Chirag S. Desai,[⊥] Kirti Shetty,[⊥] Yehia Mechref,[‡] and Habtom W. Ressom*[,†]

[†]Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC 20057, United States

[‡]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States

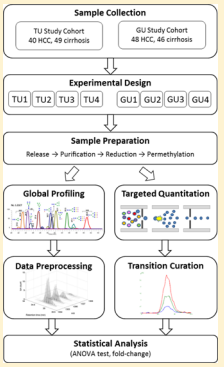
[§]Department of Mathematics and Statistics, Georgetown University, Washington, DC 20057, United States

[∥]Department of Tropical Medicine and Infectious Diseases, Tanta University, Tanta 31111, Egypt

[⊥]MedStar Georgetown University Hospital and Georgetown University Medical Center, Washington, DC 20057, United States

Supporting Information

ABSTRACT: Defining clinically relevant biomarkers for early stage hepatocellular carcinoma (HCC) in a high-risk population of cirrhotic patients has potentially far-reaching implications for disease management and patient health. Changes in glycan levels have been associated with the onset of numerous diseases including cancer. In the present study, we used liquid chromatography coupled with electrospray ionization mass spectrometry (LC−ESI-MS) to analyze N-glycans in sera from 183 participants recruited in Egypt and the U.S. and identified candidate biomarkers that distinguish HCC cases from cirrhotic controls. N-Glycans were released from serum proteins and permethylated prior to the LC−ESI-MS analysis. Through two complementary LC−ESI-MS quantitation approaches, global profiling and targeted quantitation, we identified 11 N-glycans with statistically significant differences between HCC cases and cirrhotic controls. These glycans can further be categorized into four structurally related clusters, matching closely with the implications of important glycosyltransferases in cancer progression and metastasis. The results of this study illustrate the power of the integrative approach combining complementary LC−ESI-MS based quantitation approaches to investigate changes in N-glycan levels between HCC cases and patients with liver cirrhosis.



KEYWORDS: *cancer biomarker discovery, glycomics, hepatocellular carcinoma, liver cirrhosis, mass spectrometry, multiple reaction monitoring*

INTRODUCTION

Hepatocellular carcinoma (HCC) is the third leading cause of cancer mortality worldwide with five-year relative survival rates less than 15%.[1−3] Most of the risk factors for HCC, including chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), lead to the development of liver cirrhosis, which is present in 80−90% of patients with HCC.[4] The malignant conversion of cirrhosis to HCC is often fatal in part because adequate biomarkers are not available for diagnosis during the progression stages of HCC. Survival rates of patients with HCC can be significantly improved if the diagnosis is made at earlier stages, when treatment is more effective.[3,5] Alpha-Fetoprotein (AFP), the serologic biomarker for HCC in current use, is not effective for early diagnosis due to its low sensitivity.[6,7] Therefore, more potent biomarkers for early stage HCC are needed.

Glycosylation is one of the most common post-translational modifications of proteins. Altered patterns of glycosylation have been associated with various diseases, and many currently used cancer biomarkers, including AFP, are glycoproteins.[8,9] The analysis of glycosylation is particularly relevant to liver pathology because of the major influence of this organ on the homeostasis of blood glycoproteins. Mass spectrometry is an essential tool for the analysis of glycosylation. As protein glycosylation can occur on multiple sites involving the attachment of different glycans to each site, analysis of glycoproteins requires site-specific elucidation of glycan heterogeneity.[10] This is further complicated by the different chemical properties between glycans and peptides, and analysis by mass spectrometry typically involves enrichment of glycoproteins or glycopeptides.[11] An effective alternative is to analyze glycans released from proteins and associate the glycomic changes with pathological conditions of interest. N-Glycans are of particular interest as their involvement in major biological processes, including cell−cell interactions and intracellular signaling, has important implications in disease

Special Issue: Proteomics of Human Diseases: Pathogenesis, Diagnosis, Prognosis, and Treatment

Received: May 7, 2014
Published: July 31, 2014

   dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

progression.[8] Also, several enzymes that allow efficient release of this type of glycans have been made available.[12] Through appropriate analytical methods that yield broad coverage of the glycome, characterizing glycomic patterns in serum/plasma of patients with cancer has proven a promising strategy to discover biomarkers for early diagnosis of cancer.[12−14] In particular, mass spectrometry is an enabling technology for analysis of glycans in cancer biomarker discovery.[12] The use of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) to identify N-glycan biomarkers for HCC has been widely applied and discussed.[15−18] With recent advances in mass spectrometry and separation methods, liquid chromatography−mass spectrometry (LC−MS) is capable of profiling hundreds of glycans including isomeric glycoforms.[19,20] Mass spectrometry using electrospray ionization (ESI-MS) is especially well-suited to the LC−MS based glycomic analysis. Higher sensitivity of LC−ESI-MS over MALDI-MS and LC−MALDI-MS in detecting permethylated N-glycans derived from serum has been demonstrated in a recent study.[19] However, to date glycomic profiling using LC−ESI-MS has not been fully exploited for large-scale biomarker discovery studies, and there is still a lack of appropriate computational tools.[14]

The present study applies LC−ESI-MS based serum glycomics for HCC biomarker discovery in patients with liver cirrhosis. Workflow of the proposed glycomic analysis is shown in Figure 1. Sera were collected from patients recruited in Egypt and the U.S. We utilized two complementary platforms to perform global profiling and targeted quantitation of N-glycans and identified candidate biomarkers that distinguish HCC cases from cirrhotic controls. Global profiling was performed using a high-resolution mass spectrometer (LTQ Orbitrap Velos), while targeted quantitation was performed using a triple quadrupole (QqQ) mass spectrometer in multiple reaction monitoring (MRM) mode.[21] The integrative workflow consisting of global profiling and targeted quantitation is widely applied in LC−MS based proteomic studies but to our best knowledge has not yet been exploited in glycomics. This study revealed 26 N-glycans with statistically significant differences between HCC cases and cirrhotic controls through global profiling and 15 through targeted quantitation. Eleven of these candidate N-glycan biomarkers were identified by both quantitation approaches and match closely with the implications of important glycosyltransferases in cancer progression and metastasis. The results of this study illustrate the power of the integrative approach combining LC−ESI-MS based global profiling and targeted quantitation for a comprehensive serum glycomic analysis to investigate changes in N-glycan levels between HCC cases and patients with liver cirrhosis.

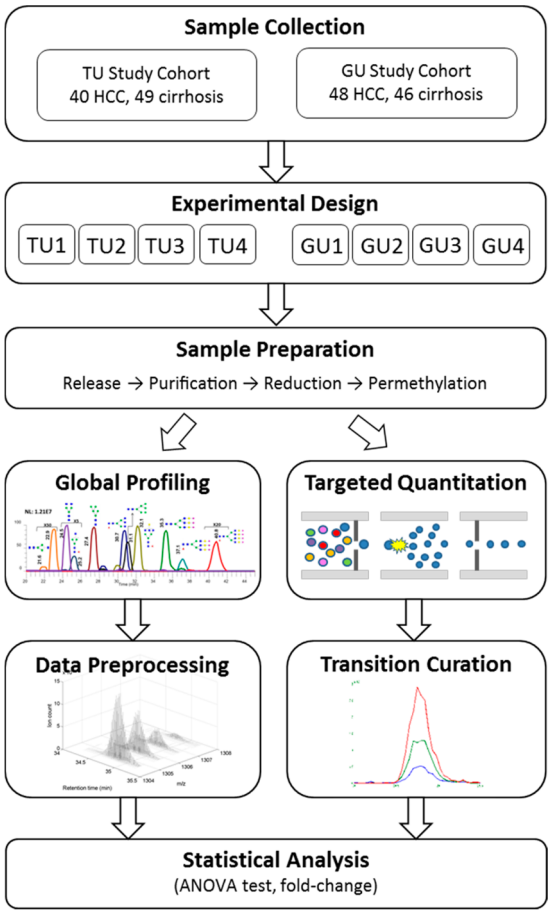


Figure 1. Workflow for the LC−ESI-MS analysis of N-glycans in sera from patients in two study cohorts (TU and GU).

STUDY COHORTS AND SAMPLE COLLECTION

The samples in this study were obtained from participants recruited in Egypt and the U.S. The Egyptian participants consisted of 89 adult patients (40 HCC cases and 49 patients with liver cirrhosis) recruited from the outpatient clinics and inpatient wards of the Tanta University Hospital, Tanta, Egypt (TU cohort). The U.S. participants comprised 94 adult patients (48 HCC cases and 46 patients with liver cirrhosis), recruited from the hepatology clinics at MedStar Georgetown University Hospital, Washington, DC, USA (GU cohort). The protocols were approved by the respective Institutional Review Boards at Tanta University and Georgetown University. Detailed characteristics of the participants in both study cohorts are provided in Supplemental Tables S1 and S2. Through peripheral venepuncture, a single blood sample was drawn into a 10 mL BD Vacutainer sterile vacuum tube without the presence of anticoagulant. The blood was immediately centrifuged at 1000*g* for 10 min at room temperature. The serum supernatant was carefully collected and centrifuged at 2500*g* for 10 min at room temperature. After aliquoting, serum was kept frozen at −80 °C until use. Primary tubes and serum aliquots were labeled using anonymous confidential code numbers with no personal identifiers. Identification codes were cross-referenced with clinical information in a pass code protected computer system. Samples were transported in dried ice.

MATERIALS AND METHODS

Experimental Design

We analyzed the collected sera in four batches for each cohort (designated as TU1, TU2, TU3, and TU4 in the TU cohort; GU1, GU2, GU3, and GU4 in the GU cohort). Each batch consisted of approximately 24 samples, balanced between HCC cases and cirrhotic controls in terms of age, race, gender, smoking, alcohol, and BMI. Samples within the same batch were prepared together, and LC−ESI-MS analysis was performed following a randomized order to avoid systematic biases (Supplemental Table S3). All of the samples were analyzed through global profiling and targeted quantitation except for three exhausted samples that were excluded from the latter. The

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

samples that were excluded during the targeted quantitation were from batches TU1, TU3, and GU2.

Chemicals and Reagents

HPLC-grade water and sodium hydroxide were obtained from Mallinckrodt Chemicals (Phillipsburg, NJ). HPLC-grade methanol, isopropyl alcohol, and acetic acid were procured from Fisher Scientific (Pittsburgh, PA). Acetonitrile (ACN) was acquired from JT Baker (Phillipsburg, NJ). Borane-ammonia complex, sodium hydroxide beads, ammonium bicarbonate, iodomethane, trifluoroacetic acid (TFA), dimethyl sulfoxide (DMSO), and formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO). Endoglycosidase purified from *Flavobacterium meningosepticum* (PNGase F, 500,000 units/mL) was obtained from New England Biolabs (Ipswich, MA).

Sample Preparation

We applied the same procedure of sample preparation in both global profiling and targeted quantitation of N-glycans in serum. The procedure includes release, purification, reduction, and permethylation of N-glycans as we have recently described.[19,20,22,23] Briefly, a 10-μL aliquot of each serum sample was mixed with 10 μL of digestion buffer (20 mM ammonium bicarbonate) and denatured in an 80 °C water bath for 1 h. Then, 1.2 μL of PNGase F (10× diluted) was added to the sample mixture. The enzymatic release was accomplished in a 37 °C water bath for 18 h. After that, a dialysis purification step was performed for 18 h to remove impurities, using an in-house-made 12-well drop-dialysis and a cellulose ester dialysis membrane with molecular weight cut-offs of 500 and 1000 Da (Spectrum Laboratories, Rancho Dominguez, CA). This was followed by the reduction of N-glycans as previously described.[19,20,22,23] A 10-μL aliquot of a borane-ammonium complex solution (10 μg/μL) was added to each sample vial and incubated at 60 °C for 1 h. Methanol was then added to the sample and dried under vacuum. Finally, we performed solid-phase permethylation as previously described.[24−26] Briefly, the dried sample was mixed with 30 μL of DMSO, 1.2 μL of water, and 20 μL of iodomethane and applied to a spin column filled with sodium hydroxide beads. After incubation for 30 min, the sample solution was centrifuged, and 20 μL of iodomethane was added. The mixture was placed back in the columns packed with sodium hydroxide beads and incubated for 20 min. Permethylated glycans were eluted out with 50 μL of ACN and dried under vacuum.

LC−ESI-MS Data Acquisition by Global Profiling

After sample preparation, the samples were continuously analyzed by LC−ESI-MS. Permethylated N-glycans were separated by an ultimate 3000 nano-LC system (Dionex, Sunnyvale, CA) with an Acclaim PepMap C_{18} column (75 μm × 15 cm, 2 μm, 100 Å) at 55 °C to prompt efficient separation. The flow rate of the nanopump was set to 350 nL/min. Mobile phase A consisted of 2% ACN and 98% water with 0.1% formic acid, while mobile phase B consisted of ACN with 0.1% formic acid. The gradient program started at 20% mobile phase B over 10 min, which was ramped to 38% at 11 min and linearly increased to 60% in the following 32 min. Then, mobile phase B was increased to 90% in 3 min, and the percentage was maintained for 4 min. Finally, mobile phase B was decreased to 20% in 1 min, and the percentage was maintained for 9 min to equilibrate the column. The nano-LC system was interfaced to an LTQ Orbitrap Velos (Thermo Scientific, San Jose, CA) hybrid mass spectrometer. The mass spectrometer was operated in data-dependent acquisition (DDA) mode, where each MS full scan (m/z range 500−2000) was followed by five MS/MS scans of the most intense ions.

Preprocessing of LC−ESI-MS Data

The data were analyzed using a preprocessing pipeline consisting of in-house-developed algorithms and open-source software tools. The preprocessing steps include deisotoping of mass spectra, peak detection, peak alignment, and normalization. We performed the deisotoping of mass spectra using DeconTools (v1.0.4672, October 16, 2012),[27] where the monoisotopic mass and charge state were deduced. DeconTools allows us to specify an appropriate average residue composition for the calculation of isotopic distribution. The average composition for the monosaccharides ($C_{10}H_{18}N_{0.43}O_5S_0$) was determined on the basis of the permethylated N-glycans commonly found in our previous studies. After the deisotoping step, peak detection was performed using an in-house-developed algorithm. Briefly, deisotoped ions with the same molecular weight (with 10 ppm tolerance) were linked along scans to generate a chromatographic trace. Low-quality traces were screened out according to user-defined criterions (i.e., minimum scans of 20 to define a peak, minimum summed intensity of 100,000, minimum density of 0.3 for valid scans in a trace, and allowable missing values of 35 between adjacent scans). Missing values in the remaining traces were interpolated using corresponding extracted ion chromatograms from raw data. The interpolated trace was further processed through successive convolution with a Savitzky-Golay smoothing filter (order of 5 and half of trace length as the window width) and a first-order derivative of a Gaussian kernel (window width of 30 scans, standard deviation of 3) to identify the position and boundary of the chromatographic peak at zero-crossing and enclosing local extrema, respectively. A detected peak was characterized by the following properties: monoisotopic mass, charge state, intensity (area under curve within boundary), and retention time. A normalization step was then applied to ensure that the summed intensity of detected peaks was identical in all of the LC−ESI-MS runs from the same batch. Peaks detected in multiple runs were aligned and matched using the simultaneous multiple alignment (SIMA, version of 2010) model[28] with the following parameters: -R 50 -M 0.1. The resulting peak list of the LC−ESI-MS runs was further refined such that only the peaks detected in over half of the runs in either case or control group were retained. Finally, missing values owing to either peak detection or alignment were interpolated using their corresponding extracted ion chromatograms. The preprocessing pipeline resulted in a consensus peak list of the LC−ESI-MS runs for subsequent analysis.

LC−ESI-MRM-MS Data Acquisition

Targeted quantitation of 117 N-glycans including isomers was performed by MRM using a QqQ mass spectrometer. These targets include (i) N-glycans that were detected on the Orbitrap or QqQ instrument in our previous studies, (ii) N-glycans evaluated as potential HCC biomarkers in previous studies,[15−18,29−31] and (iii) N-glycans involved in Golgi apparatus retrieved from KEGG GLYCAN database.[32] The 117 N-glycans were represented by 213 channels (three transitions in each) consisting of their different adduct forms and charge states. A complete list of the MRM transitions used in this study is provided in Supplemental Table S4. The MRM quantitation was performed on a TSQ Vantage mass spectrometer (Thermo Scientific, Santa Clara, CA), with Q1 and Q3 operated at a unit resolution. The chromatographic condition was as described above in the analysis of global profiling utilizing an ultimate 3000

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

Table 1. N-Glycan Candidate Biomarkers Identified by Global Profiling and Targeted Quantitation[a]

monosaccharide composition	cohort	quantitation approach	RT (min)	adduct: z (no. $[H]^+ \rightarrow [NH4]^+$)	p-value	fold change
[4-3-1-0-0]	GU	LC−ESI-MS	23.3	2 (0)	0.042	↑1.7
		MRM	26.0	2 (0)	0.018	↑1.5
[4-3-1-1-0]	GU	LC−ESI-MS	25.5	2 (0), 2 (1), 2 (2), 3 (0)	0.021, 0.001, 0.0002	↑1.4−1.7
		MRM	28.5	2 (0)	0.021	↑1.4
[4-3-2-0-0]	GU	LC−ESI-MS	24.8	2 (0)	0.009	↑1.6
		MRM	28.0	2 (0)	0.021	↑1.4
[5-3-3-0-2]	TU	LC−ESI-MS	33.0	3 (0), 4 (0)	0.005, 0.048	↑1.6
			34.6	3 (0), 4 (0)	0.022, 0.004	↑1.3−1.8
		MRM	34.3	3 (0)	0.003	↑1.5
[5-3-3-0-3]	TU	LC−ESI-MS	34.8	4 (0)	0.033	↑1.4
			36.7	3 (0)	0.007	↑1.6
		MRM	36.5	3 (0), 4 (0)	0.010, 0.029	↑1.4
[6-3-4-0-2]	TU	LC−ESI-MS	35.7	3 (0), 4 (0)	0.042, 0.030	↑1.4−1.6
		MRM	38.3	3 (0), 4 (0)	0.032, 0.024	↑1.3−1.4
[6-3-4-0-3]	TU	LC−ESI-MS	37.7	4 (0)	0.023	↑1.7
		MRM	39.5	4 (0)	0.015	↑1.5
[6-3-4-0-4]	TU	LC−ESI-MS	38.8	3 (0), 4 (0)	0.014, 0.032	↑1.5−1.6
			39.3	3 (0), 4 (0)	0.021, 0.048	↑1.5−1.7
			39.9	4 (0)	0.001	↑1.9
		MRM	36.5	4 (0)	0.018	↑1.4
			40.8	4 (0)	0.024	↑1.7
[5-3-0-0-0]	TU	LC−ESI-MS	27.6	2 (0), 2 (1), 3 (0)	0.002, 0.003, 0.011	↓1.7
		MRM	29.5	2 (0)	0.0009	↓1.8
	GU	LC−ESI-MS	25.5	2 (0), 3 (0)	0.002, 0.014	↓1.1−1.2
[5-3-1-0-0]	TU	LC−ESI-MS	29.0	2 (0), 2 (1), 3 (0)	0.008, 0.009, 0.007	↓1.3
		MRM	30.3	2 (0), 3 (0)	0.018, 0.020	↓1.3
[5-3-1-0-1]	TU	LC−ESI-MS	31.8	2 (0), 2(1)	0.026, 0.008	↓1.5−1.8
		MRM	33.8	2 (0)	0.003	↓1.4
[3-4-1-0-0]	GU	MRM	29.5	2 (0)	0.041	↑1.3
[5-3-3-2-1]	TU	MRM	37.0	3 (0), 3 (0), 3 (0)	0.025, 0.027, 0.019	↑1.3−1.4
[6-6-0-1-2]	TU	MRM	39.5	3 (1)	0.041	↑1.3
[5-3-1-1-1]	TU	MRM	36.0	2 (0)	0.027	↓1.4
	GU	MRM	35.8	2 (0)	0.002	↓1.5
[3-3-0-0-1]	TU	LC−ESI-MS	24.5	2 (1)	0.020	↑1.4
[3-3-1-1-0]	GU	LC−ESI-MS	24.3	2 (0)	0.020	↑1.7
[4-3-2-1-0]	TU	LC−ESI-MS	30.1	2 (0)	0.006	↑4.8
	GU	LC−ESI-MS	27.1	2 (0), 2 (1)	0.012, 0.005	↑1.4−1.5
			28.0	2 (0)	0.041	↑1.4
[4-3-2-0-2]	TU	LC−ESI-MS	40.2	3 (0)	0.038	↑1.4
[5-3-3-0-1]	TU	LC−ESI-MS	30.9	3 (0)	0.022	↑1.3
			32.5	3 (0)	0.006	↑1.3
[6-3-4-0-0]	TU	LC−ESI-MS	33.1	3 (0)	0.017	↑1.8
[6-3-4-0-1]	TU	LC−ESI-MS	33.7	3 (0)	0.013	↑1.3
[4-3-0-0-0]	TU	LC−ESI-MS	23.6	2 (1)	0.008	↓1.9
[4-3-0-1-0]	TU	LC−ESI-MS	25.8	2 (1)	0.009	↓1.4
	GU	LC−ESI-MS	23.3	2 (1)	0.001	↑2.3
			24.8	2 (1)	0.002	↑1.52
			36.0	3 (0)	0.040	↓1.2
[4-3-2-0-0]	TU	LC−ESI-MS	32.3	2 (0)	0.008	↓1.4
			33.3	2 (0)	0.007	↓1.5
[5-3-2-1-0]	TU	LC−ESI-MS	39.5	2 (0), 3 (0)	0.039, 0.022	↓1.2
[4-3-1-0-1]	GU	LC−ESI-MS	26.2	2 (0)	0.013	↓1.2
[4-4-2-0-2]	GU	LC−ESI-MS	33.5	3 (0)	0.017	↓1.2
[6-3-4-1-2]	GU	LC−ESI-MS	35.1	4 (0)	0.031	↓1.1
[6-3-4-1-3]	GU	LC−ESI-MS	37.0	3 (0)	0.022	↓1.1

[a]Glycans are characterized by the number of five monosaccharides: GlcNAc, mannose, galactose, fucose, and NeuNAc. The monosaccharide compositions were assigned through accurate mass matching (<2 ppm). Tandem MS spectra of the 11 glycans identified by both global profiling (LC−ESI-MS) and targeted quantitation (MRM) are provided in Figure 3. Retention times (RT, in min) in the first batch are reported. Adduct form is presented by charge state and number of protons replaced by ammonium: z (# $[H]^+ \rightarrow [NH_4]^+$). Fold change is based on the comparison of HCC versus cirrhosis, where ↑ and ↓ denote up-regulation down-regulation, respectively. Average value of fold changes in four batches is reported.

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

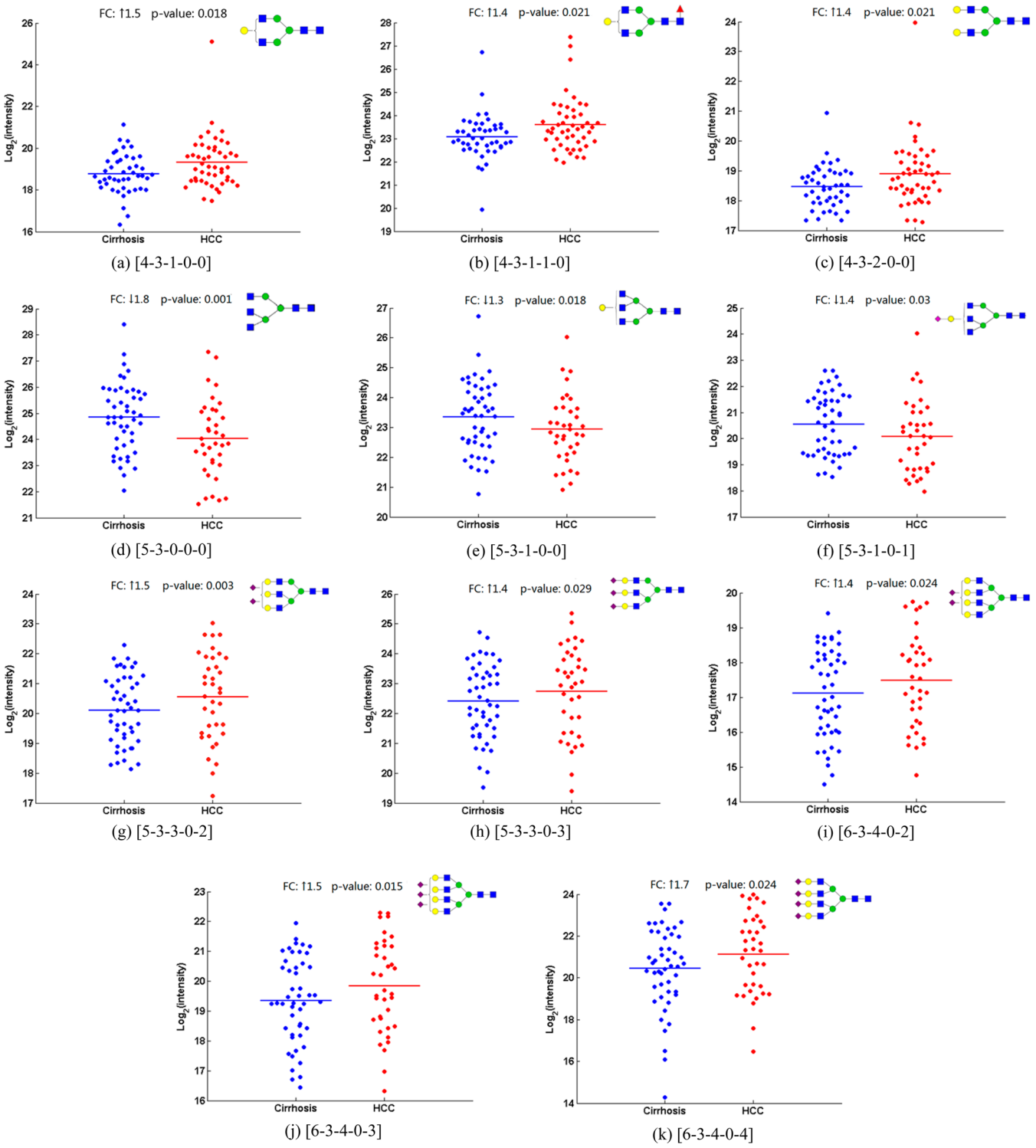


Figure 2. Quantitation results of 11 candidate N-glycan biomarkers in sera of HCC cases and cirrhotic controls by the MRM analysis. (a–c) Up-regulated biantennary glycans in the GU cohort. (d–f) Down-regulated β-1,6-GlcNAc branching glycans in the TU cohort. (g, h) Up-regulated β-1,6-GlcNAc branching glycans in the TU cohort. (i–k) Up-regulated tetra-antennary glycans in the TU cohort. FC = fold change. Blue square = GlcNAc, green circle = mannose, yellow circle = galactose, red triangle = fucose, purple diamond = NeuNAc.

nano-LC system with identical gradient setup. The cycling time of the 213 transition channels was 2.7 s on average.

Difference Detection

Following data preprocessing, the most relevant peaks with differential abundance between HCC cases and cirrhotic controls were selected using a two-way analysis of variance (ANOVA) model. Peaks from the four batches were matched upfront (m/z tolerance of 10 ppm and RT tolerance of 50 s), and peak intensity was modeled in terms of group effect (HCC versus cirrhosis), batch effect, interaction between group and batch, and random error associated with each sample. Specifically, the intensity for peak k from group i in batch j is modeled by $y_{ijk} = \mu + G_i + B_j + (G \times B)_{ij} + \epsilon_{ijk}$, where μ is the overall mean of the samples; G_i's are

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859–4868

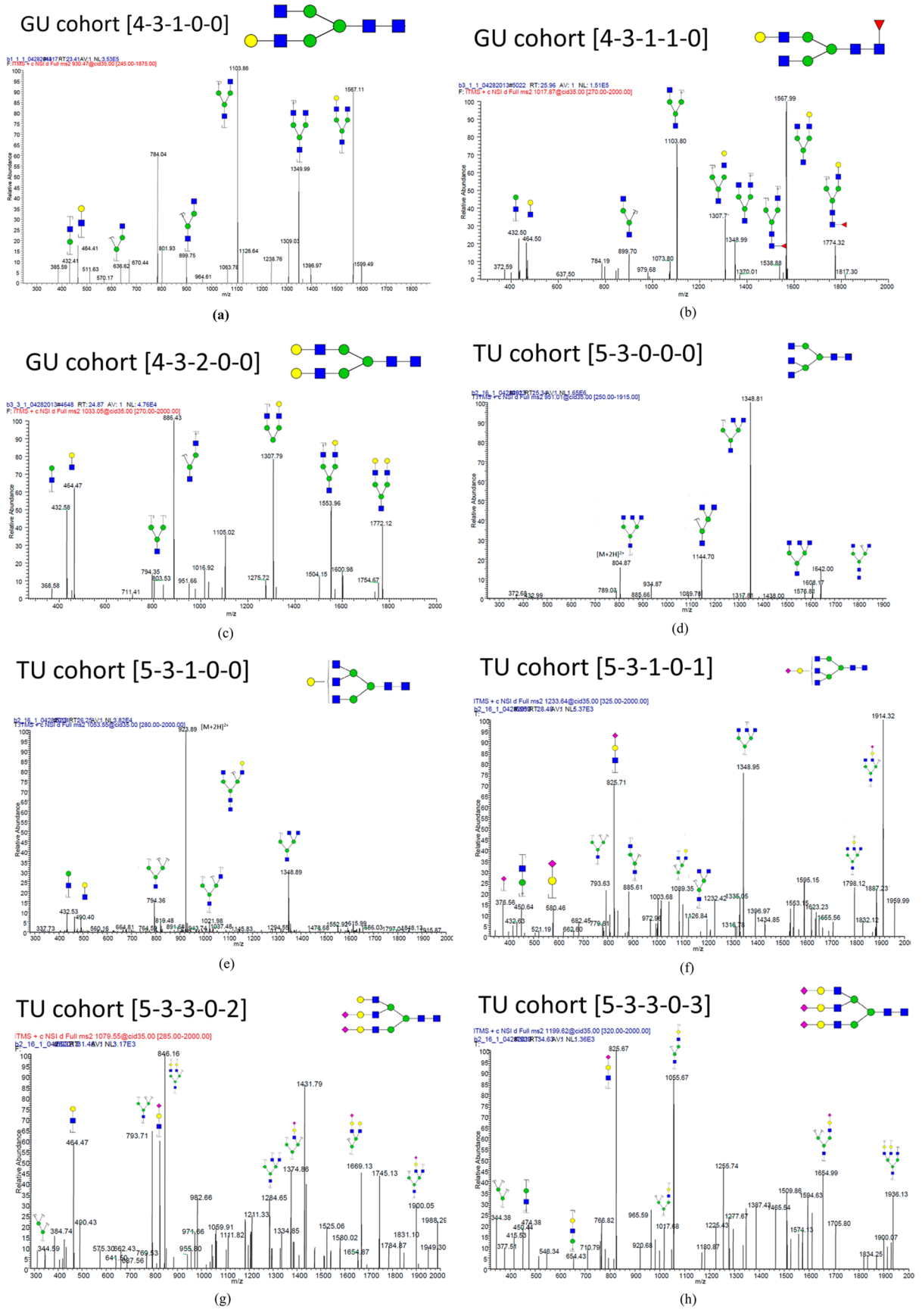


Figure 3. continued

dx.doi.org/10.1021/pr500460k | *J. Proteome Res.* 2014, 13, 4859−4868

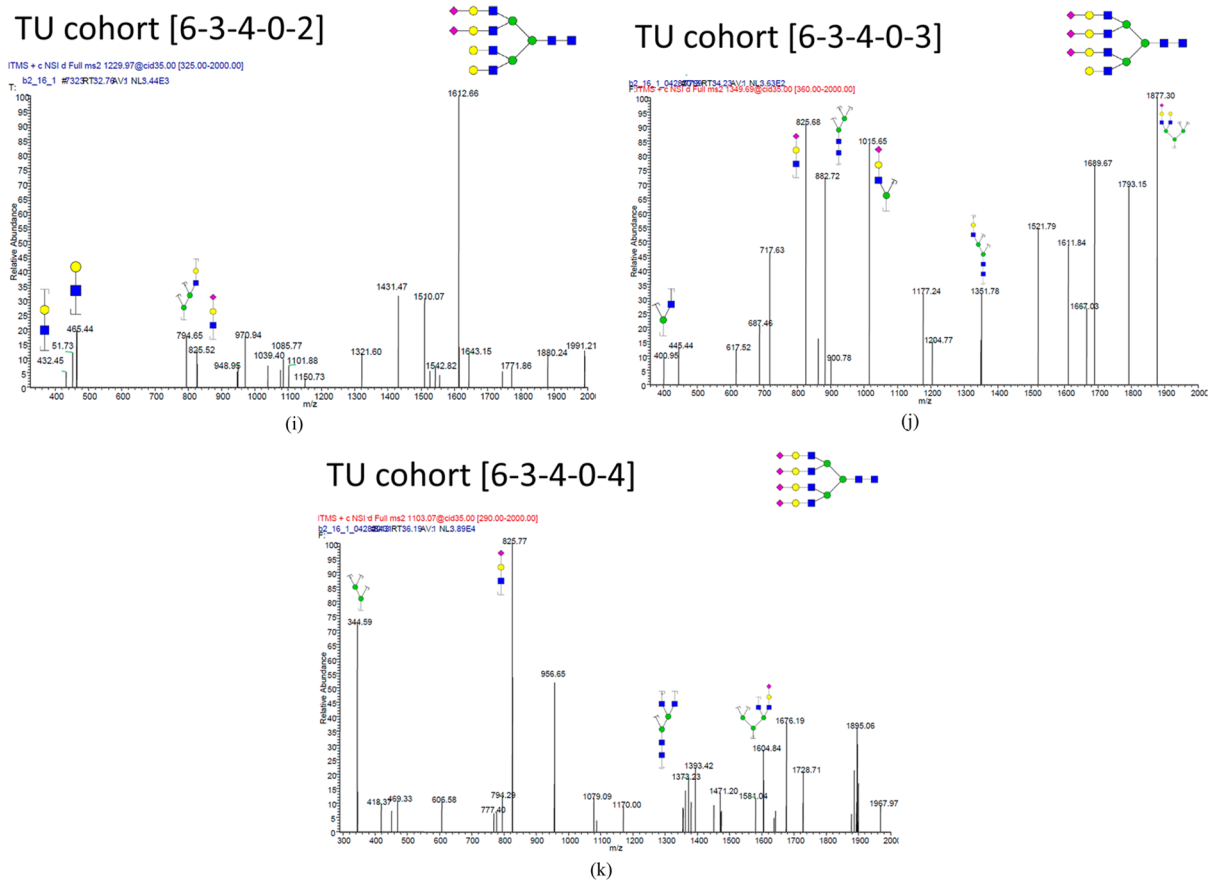


Figure 3. Annotated tandem MS spectra of 11 candidate N-glycan biomarkers: (a) [4-3-1-1-0-0], (b) [4-3-1-1-1-0], (c) [4-3-2-0-0], (d) [5-3-0-0-0], (e) [5-3-1-0-0], (f) [5-3-1-0-1], (g) [5-3-3-0-2], (h) [5-3-3-0-3], (i) [6-3-4-0-2], (j) [6-3-4-0-3], and (k) [6-3-4-0-4]. Fragment assignment was based on the criterion of signal-to-noise ratio >3.

the group effects ($\Sigma_i G_i = 0, i = 1, 2$); B_j's are the batch effects ($\Sigma_j B_j = 0, j = 1, 2, 3, 4$); $(G \times B)_{ij}$'s are the interaction between group and batch ($\Sigma_i (G \times B)_{ij} = \Sigma_j (G \times B)_{ij} = 0$); and ε_{ijk}'s are the random errors from a zero-mean normal distribution. We calculated p-values with the null hypothesis that the group means within each batch are the same. Peaks with a p-value <0.05 and having a consistent direction of fold change (FC) between groups in all four batches were selected as statistically significant.

RESULTS AND DISCUSSION

Global Profiling

LC−ESI-MS data were preprocessed in batch and peaks out of the expected RT range of glycans (15−50 min) were excluded from subsequent analysis. In the TU cohort, 2609, 3130, 2903, and 2808 peaks were detected in batches TU1, TU2, TU3, and TU4, respectively. In the GU cohort, 2559, 2519, 2556, and 2922 peaks were detected in batches GU1, GU2, GU3, and GU4, respectively. Peak lists from the four batches in each cohort were then matched. This yielded 1628 and 1500 common peaks in the TU and GU cohorts, respectively, of which 262 and 254 peaks are associated with glycans. Prior to the statistical analysis, a logarithmic transformation was applied to ensure validity of the normal distribution assumption in the ANOVA model. We evaluated the quantitation variability based on coefficient of variation (CV) of peak intensities across samples. The ranges of the CVs in the TU cohort and GU cohort are 2%−40% (with median at 8%) and 3%−58% (with median at 10%), respectively. Using the two-way ANOVA model, we found 78 peaks in the TU cohort and 91 peaks in the GU cohort that are statistically significant (p-value <0.05) and have consistent fold change across the four batches within each cohort. Putative glycan structures were assigned to the selected peaks by matching experimentally measured mass values with theoretical values of human serum N-glycans (with tolerance of 2 ppm) that were previously characterized in consideration of different charge states and adduct forms (Supplemental Table S5). Matched glycans are represented by the number of five monosaccharides: *N*-acetylglucosamine (GlcNAc), mannose, galactose, fucose, and *N*-acetylneuraminic acid (NeuNAc). This resulted in 18 significant N-glycans (11 up-regulated in HCC versus cirrhosis and 7 down-regulated) in the TU cohort and 11 significant N-glycans (6 up-regulated and 5 down-regulated) in the GU cohort (Table 1). Three glycans were found significant in both cohorts. While [4-3-2-1-0] (up-regulated) and [5-3-0-0-0] (down-regulated) have the same fold change direction in both cohorts, [4-3-0-1-0] is up-regulated in the GU cohort and down-regulated in the TU cohort.

Targeted Quantitation

Manual curation was performed to eliminate channels with unfavorable chromatographic profiles or significant noise and to determine appropriate RT windows for quantitation. Owing to the unit resolution in Q1 and Q3, interference may appear across channels with close m/z values in their transitions. The observed elution order of N-glycans on the Orbitrap system was used to elucidate some ambiguous cases in the MRM analysis. Among the 213 transition channels, 65 channels representing 82

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

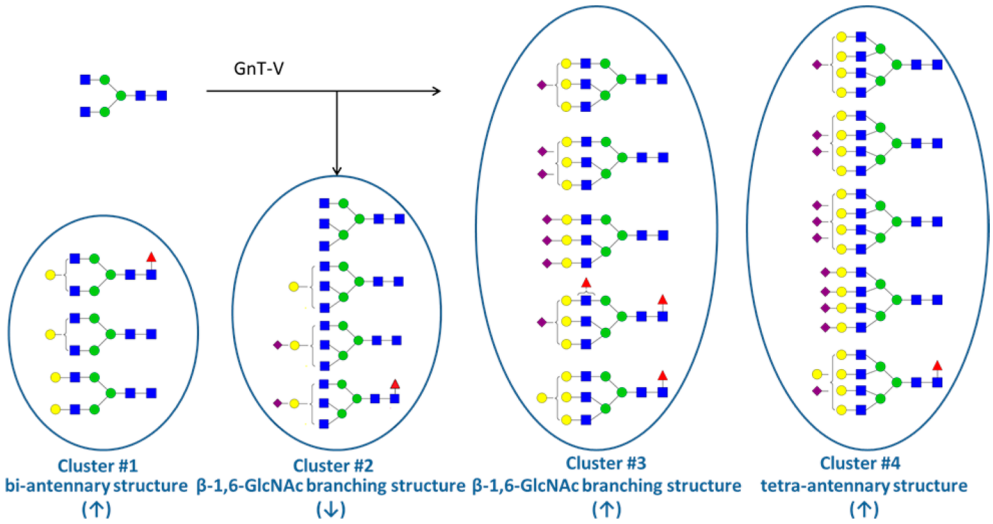


Figure 4. Four clusters of the identified N-glycan candidate biomarkers and their FC directions.

potential isomeric peaks of 52 N-glycans were detected consistently and quantitated for subsequent analysis (Supplemental Table S6). As in the global profiling analysis, peak intensities were log-transformed prior to the statistical analysis. A normalization step was also applied to ensure that the mean of the log-transformed peak intensities is identical in all the LC−ESI-MRM-MS runs from the same batch. The ranges of the CVs are 1−16% for both the TU and GU cohorts with median at 3% and 4%, respectively. Using the two-way ANOVA model, we selected significant N-glycans (p-value <0.05) and those with consistent fold changes across the four batches in each cohort. We identified 11 significant glycans (7 up-regulated and 4 down-regulated) in the TU cohort and 5 significant glycans (4 up-regulated and 1 down-regulated) in the GU cohort (Table 1). Consistent alteration (decreased level) was observed for the glycan [5-3-1-1-1] in both cohorts. Most of the significant glycans were also identified by the global profiling analysis, i.e., [4-3-1-0-0], [4-3-1-1-0], and [4-3-2-0-0] in the GU cohort and [5-3-0-0-0], [5-3-1-0-0], [5-3-1-0-1], [5-3-3-0-2], [5-3-3-0-3], [6-3-4-0-2], [6-3-4-0-3], and [6-3-4-0-4] in the TU cohort. Their MRM quantitation results and tandem MS spectra from global profiling analysis are shown in Figures 2 and 3, respectively.

In summary, we analyzed over 1500 peaks, of which over 250 are associated with glycans in global profiling. In the targeted quantitation, we monitored 82 putative isomeric peaks of 52 glycans by MRM with improved sensitivity and accuracy compared to the global profiling. Smaller CVs for quantitated glycans in MRM analysis (median CV at 4%) compared to global profiling (median CV at 10%) demonstrate the improved accuracy.

Through statistical analysis, we identified 26 and 15 significant N-glycans by global profiling and targeted quantitation, respectively. These represent 30 unique glycans, because 11 glycans overlapped between the two approaches, while the remaining 19 glycans were selected by only one of the two approaches. The latter is in part due to the stringent criterion we selected. For example, manual curation was performed before targeted quantitation to eliminate channels with unfavorable chromatographic profiles or significant noise and to keep only the reliable information. Also, only those that displayed consistent fold change direction (either up- or down-regulation) across all batches were considered. In Supplemental Table S7, we present all characteristics observed for these glycans by using the two approaches. From the table, it is noted that among the 30 reported glycans, nine are either not detected in global profiling or eliminated after curation in targeted quantitation; 11 agree between two approaches on the basis of both significance and consistent fold changes in all batches; two ([4-3-2-0-2] and [5-3-2-1-0]) have consistent fold changes in all batches but are not reported as significant in one of the two approaches; two ([5-3-1-1-1] and [4-3-2-0-0]) have consistent statistical significance but are inconsistent in terms of fold change in one of the two approaches. The remaining six glycans among 30 are not consistent on the basis of either significance or fold changes.

Most of the significant N-glycans discovered in this study are cohort-specific. This might be owing to the difference in etiologic factors between the two cohorts. Although the two cohorts involve HCC cases and cirrhotic controls, there are key differences in the characteristics of the two populations as illustrated in Supplemental Tables S1 and S2. For example, the Egyptian participants are all HCV positive and nearly all are HBV negative, whereas about half of the U.S. participants are HCV positive and about a third are HBV positive. Also, while the Egyptian participants are homogeneous Middle Eastern, the U.S. participants are approximately 56% Caucasian, 30% African American, 10% Asian, and 5% Hispanic. Moreover, about three-quarters of the Egyptian HCC cases are Stage I HCC, while Stage I HCC accounts for about half of the U.S. HCC cases. However, it should be noted that the sample size in this study is not large enough to draw solid conclusion in terms of etiology factors.

■ CONCLUSION

We analyzed N-glycans in sera from HCC cases and cirrhotic controls recruited in Egypt and the U.S. Specifically, N-glycans were enzymatically removed from serum proteins and permethylated, allowing relative quantitation of hundreds of oligosaccharides. Candidate N-glycan biomarkers were identified through LC−ESI-MS based global profiling and targeted quantitation. The most relevant glycans in distinguishing HCC cases from cirrhotic controls were selected using a two-way ANOVA model. We identified 26 and 15 statistically significant N-glycans through global profiling and targeted quantitation, respectively. Although none of these glycans had an adjusted p-value <0.05 in consideration of multiple testing correction using

dx.doi.org/10.1021/pr500460k | J. Proteome Res. 2014, 13, 4859−4868

the method by Benjamini and Hochberg,[33] our integrative analysis revealed a good overlap of the significant glycans identified by both quantitation approaches. There are 11 candidate biomarkers overlapping between the two complementary platforms: [4-3-1-0-0], [4-3-1-1-0], [4-3-2-0-0], [5-3-0-0-0], [5-3-1-0-0], [5-3-1-0-1], [5-3-3-0-2], [5-3-3-0-3], [6-3-4-0-2], [6-3-4-0-3], and [6-3-4-0-4]. Eight of these glycans are up-regulated in HCC versus cirrhosis, while three are down-regulated. The up-regulation of the bigalactose biantennary glycan [4-3-2-0-0] is in agreement with a previously reported study.[30] The overlap between our findings and previous glycomic studies is limited, because sialylated glycans were often excluded from the previous analysis.[29−31,34]

Biosynthesis of N-glycans in the Golgi apparatus involves trimming of mannose residuals and stepwise addition of monosaccharides, resulting in three groups of N-glycans: high-mannose, complex, and hybrid types. Most of the candidate biomarkers identified in this study are complex type N-glycans with two hybrid-type glycans: [3-4-1-0-0] and [6-6-0-1-2]. Some biomarker discovery studies for other types of cancer have shown that structurally related glycans are likely to have correlated changes of levels due to the same biosynthesis process they are involved in.[14] Our study exhibited a similar phenomenon, where many of the candidate biomarkers for HCC are closely related in their structures. These glycans can be grouped into four clusters: Cluster 1, biantennary structure; Cluster 2, β-1,6-GlcNAc branching structure (down-regulated); Cluster 3, β-1,6-GlcNAc branching structure (up-regulated); and Cluster 4, tetra-antennary structure as shown in Figure 4. Within each cluster the glycans show consistent changes in their levels.

Further elucidation of the relationship between the identified complex N-glycans can be obtained by referring to their biosynthesis process. Glycosyltransferases such as *N*-acetylglucosaminyltransferase V (GnT-V) have been known to play a key role in the formation of N-glycan branches.[35,36] GnT-V catalyzes the addition of β-1,6-GlcNAc branching of N-glycans and has been considered as a promoter of metastasis. Their implications on HCC have also been discussed.[30,34,37] In this study, increased levels in HCC were found in GnT-V's downstream products (Clusters 3 and 4 in Figure 4), matching the role of GnT-V in cancer metastasis. We note, however, that opposite alteration was observed in truncated branching glycans, i.e., Cluster 2 where addition of galactose to antennary is not complete. In addition, increased levels of β-galactoside α-2,6-sialyltransferase (ST6GalI), which transfers sialic acid residue in α-2,6 linkage to a terminal galactose, have been associated with progression and poor prognosis in HCC.[9,38−40] Consistent findings were obtained in this study, where increased levels in HCC were found in a number of sialylated GlcNAc branching N-glycans (downstream products of ST6GalI).

This study revealed that glycomic changes during cancer progression represent systematic alteration. Enrichment analysis could potentially increase the statistical power to detect the glycomic changes on a systems level and yield biologically relevant results, as demonstrated in the genomic analysis.[41] Comprehensive coverage of the glycome is a prerequisite, and the presented LC−ESI-MS based workflow is expected to serve as a primary approach for this type of analysis. In order to allow a rigorous integration of changes in multiple glycans, defining appropriate categories of ontological/topological information is critical. This strategy may be further enhanced by integrating additional characteristics through other -omic analysis, such as proteomics and glycoproteomics.

Our future studies will focus on investigating these candidate biomarkers on a larger population that allows stratification of the subjects on the basis of etiology and disease stage. We will also perform multivariate analysis or enrichment analysis to identify a panel of HCC biomarkers and to evaluate if the observed glycomic changes can be reliably used for early detection of HCC in high risk population of cirrhotic patients. Furthermore, we will investigate the underlying biochemical processes driving the glycomic changes and incorporate additional -omic measurements on the same subjects for a more comprehensive characterization.

■ ASSOCIATED CONTENT

Supporting Information

Characteristics of the study cohorts, ethical approvals, LC−ESI-MS data acquisition order, structure and theoretical mass of N-glycans characterized by different charge states and adduct forms, curation of MRM transitions, and a list of statistically significant glycans obtained by global profiling and targeted quantitation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 202-687-2283. Fax: 202-687-0227. E-mail: hwr@georgetown.edu.

Author Contributions

¶These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by NIH Grants R01CA143420 and R01GM086746.

■ ABBREVIATIONS

ACN, acetonitrile; AFP, α-fetoprotein; ANOVA, analysis of variance; DMSO, dimethyl sulfoxide; ESI, electrospray ionization; FA, formic acid; GlcNAc, *N*-acetylglucosamine; GnT-V, *N*-acetylglucosaminyltransferase V; HCC, hepatocellular carcinoma; HPLC, high-performance liquid chromatography; LC−MS, liquid chromatography coupled with mass spectrometry; MRM, multiple reaction monitoring; NeuNAc, *N*-acetylneuraminic acid; QqQ, triple quadrupole; RT, retention time; ST6GalI, β-galactoside α-2,6-sialyltransferase; TFA, trifluoroacetic acid

■ REFERENCES

(1) Ferlay, J.; Shin, H. R.; Bray, F.; Forman, D.; Mathers, C.; Parkin, D. M. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **2010**, *127* (12), 2893−2917.

(2) Arzumanyan, A.; Reis, H. M.; Feitelson, M. A. Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nat. Rev. Cancer* **2013**, *13* (2), 123−135.

(3) Center, M.; Siegel, R.; Jemal, A. *Global Cancer Facts & Figures*; American Cancer Society: Atlanta, GA. 2011.

(4) El-Serag, H. B. Hepatocellular carcinoma. *N. Engl. J. Med.* **2011**, *365* (12), 1118−1127.

(5) Bialecki, E. S.; Di Bisceglie, A. M. Diagnosis of hepatocellular carcinoma. *HPB (Oxford)* **2005**, *7* (1), 26−34.

(6) Trevisani, F.; D'Intino, P. E.; Morselli-Labate, A. M.; Mazzella, G.; Accogli, E.; Caraceni, P.; Domenicali, M.; De Notariis, S.; Roda, E.; Bernardi, M. Serum alpha-fetoprotein for diagnosis of hepatocellular

carcinoma in patients with chronic liver disease: Influence of HBsAg and anti-HCV status. *J. Hepatol.* **2001**, *34* (4), 570−575.
(7) Gupta, S.; Bent, S.; Kohlwes, J. Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann. Int. Med.* **2003**, *139* (1), 46−50.
(8) Fuster, M. M.; Esko, J. D. The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nat. Rev. Cancer* **2005**, *5* (7), 526−542.
(9) Blomme, B.; Van Steenkiste, C.; Callewaert, N.; Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *J. Hepatol.* **2009**, *50* (3), 592−603.
(10) An, H. J.; Froehlich, J. W.; Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 421−426.
(11) Zaia, J. Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **2008**, *15* (9), 881−892.
(12) Mechref, Y.; Hu, Y.; Garcia, A.; Hussein, A. Identifying cancer biomarkers by mass spectrometry-based glycomics. *Electrophoresis* **2012**, *33* (12), 1755−1767.
(13) An, H. J.; Kronewitter, S. R.; de Leoz, M. L.; Lebrilla, C. B. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **2009**, *13* (5−6), 601−607.
(14) Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol. Cell. Proteomics* **2013**, *12* (4), 846−855.
(15) Ressom, H. W.; Varghese, R. S.; Goldman, L.; Loffredo, C. A.; Abdel-Hamid, M.; Kyselova, Z.; Mechref, Y.; Novotny, M.; Goldman, R. Analysis of MALDI-TOF mass spectrometry data for detection of glycan biomarkers. *Pac. Symp. Biocomput.* **2008**, 216−227.
(16) Tang, Z.; Varghese, R. S.; Bekesova, S.; Loffredo, C. A.; Hamid, M. A.; Kyselova, Z.; Mechref, Y.; Novotny, M. V.; Goldman, R.; Ressom, H. W. Identification of N-glycan serum markers associated with hepatocellular carcinoma from mass spectrometry data. *J. Proteome Res.* **2010**, *9* (1), 104−112.
(17) Goldman, R.; Ressom, H. W.; Varghese, R. S.; Goldman, L.; Bascug, G.; Loffredo, C. A.; Abdel-Hamid, M.; Gouda, I.; Ezzat, S.; Kyselova, Z.; Mechref, Y.; Novotny, M. V. Detection of hepatocellular carcinoma using glycomic analysis. *Clin. Cancer Res.* **2009**, *15* (5), 1808−1813.
(18) Kamiyama, T.; Yokoo, H.; Furukawa, J.; Kurogochi, M.; Togashi, T.; Miura, N.; Nakanishi, K.; Kamachi, H.; Kakisaka, T.; Tsuruga, Y.; Fujiyoshi, M.; Taketomi, A.; Nishimura, S.; Todo, S. Identification of novel serum biomarkers of hepatocellular carcinoma using glycomic analysis. *Hepatology* **2013**, *57* (6), 2314−2325.
(19) Hu, Y.; Mechref, Y. Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC−ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum. *Electrophoresis* **2012**, *33* (12), 1768−1777.
(20) Desantos-Garcia, J. L.; Khalil, S. I.; Hussein, A.; Hu, Y.; Mechref, Y. Enhanced sensitivity of LC−MS analysis of permethylated N-glycans through online purification. *Electrophoresis* **2011**, *32* (24), 3516−3525.
(21) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **2010**, *7* (1), 43−46.
(22) Hu, Y.; Desantos-Garcia, J. L.; Mechref, Y. Comparative glycomic profiling of isotopically permethylated N-glycans by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2013**, *27* (8), 865−877.
(23) Huang, Y.; Mechref, Y.; Novotny, M. V. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. *Anal. Chem.* **2001**, *73* (24), 6063−6069.
(24) Kang, P.; Mechref, Y.; Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22* (5), 721−734.
(25) Mechref, Y.; Kang, P.; Novotny, M. V. Solid-phase permethylation for glycomic analysis. *Methods Mol. Biol.* **2009**, 53−64.
(26) Kang, P.; Mechref, Y.; Klouckova, I.; Novotny, M. V. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom.* **2005**, *19* (23), 3421−3428.
(27) Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinf.* **2009**, DOI: 10.1186/1471-2105-10-87.
(28) Voss, B.; Hanselmann, M.; Renard, B. Y.; Lindner, M. S.; Kothe, U.; Kirchner, M.; Hamprecht, F. SIMA: Simultaneous multiple alignment of LC/MS peak lists. A. *Bioinformatics* **2011**, *27* (7), 987−993.
(29) Tanabe, K.; Deguchi, A.; Higashi, M.; Usuki, H.; Suzuki, Y.; Uchimura, Y.; Kuriyama, S.; Ikenaka, K. Outer arm fucosylation of N-glycans increases in sera of hepatocellular carcinoma patients. *Biochem. Biophys. Res. Commun.* **2008**, *374* (2), 219−225.
(30) Liu, X. E.; Desmyter, L.; Gao, C. F.; Laroy, W.; Dewaele, S.; Vanhooren, V.; Wang, L.; Zhuang, H.; Callewaert, N.; Libert, C.; Contreras, R.; Chen, C. N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. *Hepatology* **2007**, *46* (5), 1426−1435.
(31) Debruyne, E. N.; Vanderschaeghe, D.; Van Vlierberghe, H.; Vanhecke, A.; Callewaert, N.; Delanghe, J. R. Diagnostic value of the hemopexin N-glycan profile in hepatocellular carcinoma patients. *Clin. Chem.* **2010**, *56* (5), 823−831.
(32) Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K. F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; Kanehisa, M. KEGG as a glycome informatics resource. *Glycobiology* **2006**, *16* (5), 63R−70R.
(33) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995**, *57* (1), 289−300.
(34) Mehta, A.; Norton, P.; Liang, H.; Comunale, M. A.; Wang, M.; Rodemich-Betesh, L.; Koszycki, A.; Noda, K.; Miyoshi, E.; Block, T. Increased levels of tetra-antennary N-linked glycan but not core fucosylation are associated with hepatocellular carcinoma tissue. *Cancer Epidemiol. Biomarkers Prev.* **2012**, *21* (6), 925−933.
(35) Zhao, Y.; Sato, Y.; Isaji, T.; Fukuda, T.; Matsumoto, A.; Miyoshi, E.; Gu, J.; Taniguchi, N. Branched N-glycans regulate the biological functions of integrins and cadherins. *FEBS J.* **2008**, *275* (9), 1939−1948.
(36) Pinho, S. S.; Reis, C. A.; Paredes, J.; Magalhaes, A. M.; Ferreira, A. C.; Figueiredo, J.; Xiaogang, W.; Carneiro, F.; Gartner, F.; Seruca, R. The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Hum. Mol. Genet.* **2009**, *18* (14), 2599−2608.
(37) Yao, M.; Zhou, D. P.; Jiang, S. M.; Wang, Q. H.; Zhou, X. D.; Tang, Z. Y.; Gu, J. X. Elevated activity of N-acetylglucosaminyltransferase V in human hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.* **1998**, *124* (1), 27−30.
(38) Zhao, Y.; Li, Y.; Ma, H.; Dong, W.; Zhou, H.; Song, X.; Zhang, J.; Jia, L. Modification of sialylation mediates the invasive properties and chemosensitivity of human hepatocellular carcinoma. *Mol. Cell. Proteomics* **2014**, *13* (2), 520−536.
(39) Pousset, D.; Piller, V.; Bureaud, N.; Monsigny, M.; Piller, F. Increased alpha2,6 sialylation of N-glycans in a transgenic mouse model of hepatocellular carcinoma. *Cancer Res.* **1997**, *57* (19), 4249−4256.
(40) Hedlund, M.; Ng, E.; Varki, A.; Varki, N. M. Alpha 2−6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. *Cancer Res.* **2008**, *68* (2), 388−394.
(41) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (43), 15545−15550.