Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PRM-MS Quantitative Analysis of Isomeric *N*-Glycopeptides Derived From Human Serum Haptoglobin of Patients With Cirrhosis and Hepatocellular Carcinoma

Cristian D. Gutierrez Reyes
Texas Tech University
Yifan Huang
Texas Tech University
Mojgan Atashi
Texas Tech University
Jie Zhang
University of Michigan–Ann Arbor
Jianhui Zhu
University of Michigan–Ann Arbor
Suyu Liu
The University of Texas MD Anderson Cancer Center
Neehar D. Parikh
University of Michigan–Ann Arbor
Amit G. Singal
The University of Texas Southwestern Medical Center
J. Dai
The University of Texas MD Anderson Cancer Center
David M. Lubman
University of Michigan–Ann Arbor
Yehia Mechref (yehia.mechref@ttu.edu)
Texas Tech University

Research Article

Keywords: PRM, N-Glycopeptides, Haptoglobin, Cirrhosis, Hepatocellular Carcinoma (HCC)

Posted Date: April 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-399869/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Currently surveillance strategies have inadequate performance for the early detection of hepatocellular carcinoma (HCC). Protein glycosylation is a potential source of biomarkers to differentiate between cirrhosis and HCC. We performed a comprehensive LC-PRM-MS approach where a targeted parallel reaction monitoring (PRM) strategy was coupled to a powerful LC system to study the microheterogeneity of haptoglobin (Hp) extracted from 15 patients with cirrhosis and 15 with HCC. We found that our strategy was able to identify a large number of isomeric *N*-glycopeptides mainly located in the glycosylation site Asn207. Nine out of twelve *N*-glycopeptides, located in the Asn207 site, had significant differences in abundance between patients with cirrhosis and HCC ($p < 0.05$). The area under the curve (AUC) of alpha-fetoprotein (AFP) was alone 0.85, which improved to 0.95 (95% CI: 0.88, 1) when NLF_5613 Isomer 1 was combined with AFP. When comparing the early HCC vs. cirrhosis, four sialylated-fucosylated glycopeptides better estimated AUCs with respect to AFP (AUC_{AFP} = 0.66, and $AUC_{N\text{-glycopeptides}}$ = 0.86, 0.84, 0.88, and 0.80, respectively). Further large scale validation of glycopeptides for the early detection of HCC is warranted.

Introduction

Glycosylation alterations of serum proteins are commonly associated with the development of several cancer types; therefore, changes in protein glycosylation may play an important role in disease diagnosis. [1–3] The development of accurate biomarkers for early stages of high mortality rate cancer types such as hepatocellular carcinoma (HCC) [4] may provide opportunities for more effective patient treatments [5]. Most HCCs develops in the setting of cirrhosis [6], however current early detection strategies, using abdominal ultrasound and alpha-fetoprotein (AFP), have inadequate sensitivity and specificity for early detection [7]. The performance of ultrasound is further diminished in patients with nonalcoholic steatohepatitis (NASH) associated with central obesity, and the fastest growing cause of HCC in the US [8]. An accurate biomarker based strategy, particularly in patients with NASH, has the potential to improve HCC related outcomes.

In recent years, glycomic studies of serum haptoglobin have identified important isomeric structures, most of them sialylated fucosylated glycans with significant differences in abundance between patients with cirrhosis and HCC. [1,2] Analyses of haptoglobin *N*-glycopeptides have described the heterogeneity of its glycosylation sites, and have identified potential sialylated fucosylated glycopeptide structures as potential biomarker candidates for HCC [9–11]. According to these studies, glycoproteomic analysis focused on an extensive determination of isomeric *N*-glycopeptide structures could increase the possibility of finding biomarkers that can detect early stage HCC. Moreover, the glycosylation site information obtained by glycoproteomic analysis would be an important factor in understanding how aberrant glycosylation is related to disease progression [12].

The structural identification and quantitation of isomeric site-specific glycosylation of haptoglobin remains challenging due to the low abundance of important sialylated fucosylated structures [1,3,9,13]. In

order to address an accurate LC-MS quantitation of low abundance *N*-glycopeptide structures, the MS sensitivity needs to be enhanced and the LC separation improved. Current strategies to increase glycopeptide sensitivity use either enrichment or fractionation methods [14–16]. Examples include the immunoprecipitation of glycoproteins from complex serum and plasma samples, and the separation of sialylated glycopeptides based on their structural differences using HILIC or lectin enrichment approaches [17–20].

An MS-based approach that can provide enhanced sensitivity is multiple reaction monitoring (MRM), which increases the glycopeptide detection by scanning the specific ions that represent the target structures. MRM increases the analyte detection sensitivity but is limited by its inability to perform MS^n identification of the scanned ions, which is an important drawback in the evaluation of complex biological samples [21,22]. Alternatively, parallel reaction monitoring (PRM) is a promising strategy for identifying and quantifying specific glycopeptides in complex biological samples. This technique is possible due to modern highly sensitive mass analyzers such as the Orbitrap and Time-of-Flight (TOF) that are capable of collecting full MS^n spectra of the targeted ions, where the analytes can be accurately identified based on their *m/z* ratio and fragmentation patterns, and quantified using their most abundant and constant fragments [15,21]. For PRM quantitation the fragmentation of the targeted ion is a relevant parameter that needs to be accurately evaluated, as most fragmentation mechanisms affect the glycan and peptide moieties of the glycopeptide differently. A soft fragmentation strategy such as collision induced dissociation (CID) provides high abundance and constant glycopeptide B and Y ions from the glycan portion, commonly glycan fragments without peptide information [23]. In contrast, high energy collision dissociation (HCD) and electron transfer/higher-energy collision dissociation (EThcD) strategies generate glycan and peptide fragments which allow for complete identification of the glycopeptide structures [9,15,24].

In this study, we integrated two strategies to evaluate the microheterogeneity of serum haptoglobin focusing on the glycosylation sites Asn184, Asn207, and Asn241. Haptoglobin was extracted from serum samples from patients with NASH cirrhosis and NASH related-HCC. The first strategy involved the use of a long nano-C18 column (50 cm) to ensure a wide separation of the tryptic and Glu-C digested haptoglobin isomeric *N*-glycopeptides. The second strategy was based on a MS-PRM approach directed to enhance the detection sensitivity of important low abundance sialylated and fucosylated glycopeptide structures that are related to the presence of HCC [13,25]. We were able to identify site-specific glycosylation changes in serum haptoglobin extracted from patients with NASH cirrhosis and NASH related-HCC, where these changes were strongly related to the development of *N*-glycopeptide isomeric structures. Of significance, the statistical analysis of the data found twelve *N*-glycopeptide structures that can differentiate between both cohorts, therefore identifying *N*-glycopeptides with potential use as biomarkers for early detection of HCC in patients with cirrhosis.

Results And Discussion

A targeted PRM approach was applied for the quantitation of intact *N*-glycopeptides of serum haptoglobin to characterize the alterations in site-specific glycopeptide forms between cirrhosis and HCC.

The experimental workflow is shown in Fig. 1. Haptoglobin was purified from a 20µL aliquot of serum from each patient by using an HPLC-based antibody-immobilized column [1,9]. Serum haptoglobin has four glycosylation sites (Asn184, Asn207, Asn211, and Asn241), which can be observed after a two-step enzymatic digestion using trypsin and Glu-C [9]. In this work we focused on the glycosylation sites Asn184, Asn207, and Asn241; where we were unable to identify any *N*-glycopeptides from the glycosylation site Asn211 in a reproducible number of samples to apply reliable evaluation [9,11]. A pooled sample was analyzed and used to identify the *N*-glycopeptides present in the samples based on monoisotopic mass, charge, retention time, and MS^2 spectra, **Supplementary Table S1**.

The specific information for each detected *N*-glycopeptide structure was targeted in the LC-MS-PRM approach on the Orbitrap QExactive (Thermo). The selections of the precursor ions for all identified glycopeptide structures was based on the signal intensity so that each case the most abundant ion was used in the PRM strategy. Three charged precursor ions were the common ionic glycopeptide species in the three evaluated glycosylation sites, except for the glycopeptides with tri- and tetra-antennary and mono-, di-, and trisialylated glycans attached to the glycosylation site Asn241; where these structures presented four charged ions, **Supplementary Table S1**. The energy level was tested in the HCD cell of the QExactive before the application of the PRM strategy. The most appropriate collision energy to produce stable and abundant Y fragment ions of the glycan portion was 25 eV. Additionally, the fragmentation produced allowed us to confirm the site-specific glycosylation by the observation of abundant Y1 ions with *m/z* values of 1940.9333, 1176.5484, and 1998.0920 for the sites Asn184, Asn207, and Asn241 respectively. The peptide fragments were observed in low abundance in the three glycosylation sites, **Supplementary Fig. S1**. Six of the most representative and abundant fragment Y ions were selected for each *N*-glycopeptide, and their peak area calculated using Xcalibur (Thermo) software, see **Supplementary Table S1**. After the area under the curve was computed the data was normalized based on the total area of all identified glycopeptides and their relative abundance calculated. Significant differences in abundance between cirrhosis and HCC samples were calculated using the Wilcoxon Test. Any $p-$ values < 0.05 were considered to be significant changes; **Supplementary Table S2** shows the data obtained for cirrhosis samples and **Supplementary Table S3** shows the data obtained for HCC samples. The retention time of the evaluated structures was such that *N*-glycopeptides for the glycosylation site Asn184, eluted between 42 and 54 min, for the Asn207 site between 40 and 60 min, and for the Asn241 site between 61 and 88 min. For all sites a retention time pattern based on the glycan moiety was observed, **Supplementary Table S1**. For all the glycosylation sites the first glycopeptides to be eluted were the structures with small biantennary mono-, and di-sialylated glycans attached. The glycopeptides with sialylated-fucosylated glycans attached eluted according to the number of sialic acid molecules presented in the glycan moiety regardless of the antennae number, **Supplementary Table S1**. The core or branch fucosylation of important *N*-glycopeptides was performed by the evaluation of their fragmentation, an example can be found in **Supplementary Fig. S2**. A pooled sample was used to

determine the sialic acid linkage of some important structures by the application of α2,3 neuraminidase enzyme digestion (a description can be found in **Supplementary Fig. S3**). To facilitate the glycan and glycopeptide descriptions, the nomenclature used is as follows. For glycans, a four-digit nomenclature represents the number of HexNAc, Hex, Fuc, and Neu5Ac molecules (*N*-acetylhexosamine, Hexose, Fucose, and *N*-acetylneuraminic acid respectively). In the case of glycopeptides, the letters MVS, NFL, and VVL were added to the glycan nomenclature and represent the peptide chain where they were observed (MVS = MVSHHN184LTTGATLINE, NLF = NLFLN207HSE, and VVL = VVLHPN241YSQVDIGLIK).

Figure 2 describes the data treatment using the glycopeptide structure NLFLN207HSE + $HexNAc_5$,Hex_6,$Neu5Ac_3$ (NLF_5603) as an example. Figure 2a and Fig. 2b depict the extracted ion chromatograms (EICs) for the glycopeptide NLF_5603 derived from all cirrhosis and HCC patients, respectively. According to the statistical test, the isomeric structure number 2 of this glycopeptide form had significant changes in abundance between cirrhosis and HCC samples, **Supplementary Table S4**; this result is further discussed in this section. The corresponding precursor ion for the structure NLF_5603 has a triple charge *m/z* value of 1181.7977, it eluted in a retention time of about 72 min, and it produced monocharged Y fragment ions principally of the glycan portion. Figure 2c shows the capability of our strategy to resolve three isomeric structures for the glycopeptide NLF_5603. Representative TICs of other important isomeric structures observed in the samples are described in **Supplementary Fig. S4**. Figure 2d shows the sensitivity enhancement of our PRM approach versus the full scan (MS^1) signal observed in the pooled sample injected with the same concentration. This increase on sensitivity allowed us to address an accurate quantitation of low abundant sialylated fucosylated glycopeptide structures that have special importance in HCC detection.

Haptoglobin Microheterogeneity

Based on our early research efforts as described by Huang et al. and Zhu et al., the isomeric glycan profile of serum haptoglobin and its glycosylation site heterogeneity were determined [1,9]. These works demonstrated the importance of the isomeric structures, either glycan or glycopeptides in the differentiation of patients with NASH cirrhosis and NASH related-HCC. Huang et al. reported seven sialylated fucosylated isomeric glycans with the structures $HexNAc_4Hex_5FucNeuAc$, $HexNAc_5Hex_6FucNeuAc_3$, and $HexNAc_5Hex_6Fuc_2NeuAc_3$. All of these structures showed significant changes in abundance between the two disease states [1]. In a related work, Jin et al. found 12 isomeric glycans with significant differences between healthy and HCC patients [13]. The development of an LC strategy that allows the separation of isomeric *N*-glycopeptides was addressed based on the use of a long C18 column (50cm) and high temperature. This strategy was adopted from the work of Ji et al., where they reported an important separation of isomeric *O*- and *N*-glycopeptides from tryptic digested α1-acid glycoprotein (AGP), and evaluated the effects of temperature on the isomeric separation as well [26].

We were able to accurately identify and quantify a total of 73 isomeric structures that corresponded to 42 *N*-glycopeptide forms from the tryptic/Glu-C digested haptoglobin, as shown in Table 1. The distribution

of the *N*-glycopeptides in the glycosylation sites were as follows: 13 isomeric structures that corresponded to 11 *N*-glycopeptide forms in the glycosylation site Asn184; 44 isomeric structures that corresponded to 19 *N*-glycopeptide forms in the glycosylation site Asn207; and 16 isomeric structures that corresponded to 12 *N*-glycopeptide forms in the glycosylation site Asn241. The glycosylation site Asn207 was found to have the most abundant diversity of *N*-glycopeptides as well as the most abundant isomeric glycopeptide structures.

Table 1

Observed haptoglobin glycopeptide structures by glycosylation site including isomeric forms. The number in parentheses "()" represents the number of isomeric forms in the glycopeptide structure. Glycan nomenclature; HexNAc, Hex, Fuc, NeuAc (*N*-acetylhexosamine, Hexose, Fucose, *N*-acetylneuraminic acid).

Glycopeptide backbone	**MVSHHN184LTTGATLINE**	**NLFLN207HSE**	**VVLHPN241YSQVDIGLIK**
N-**glycosylation Microheterogeneity**	---	[3-4-0-1](**2**)	---
	---	[4-4-0-1](**3**)	[4-4-0-1]
	[4-5-0-0] (**2**)	[4-5-0-0](**2**)	[4-5-0-0]
	[4–5–1–1]	[4–5–1–1](**2**)	---
	[4–5–1–2]	[4–5–1–2](**2**)	[4–5–1–2](**2**)
	[4-5-0-1]	[4-5-0-1]	[4-5-0-1]
	[4-5-0-2]	[4-5-0-2](**2**)	[4-5-0-2]
	[4-6-0-1]	---	---
	[5–6–1–1]	[5–6–1–1](**2**)	---
	[5–6–1–2]	[5–6–1–2](**2**)	[5–6–1–2]
	[5–6–1–3]	[5–6–1–3](**2**)	---
	[5-6-0-1]	[5-6-0-1](**2**)	[5-6-0-1]
	[5-6-0-2](**2**)	[5-6-0-2](**2**)	[5-6-0-2]
	---	[5-6-0-3](**3**)	[5-6-0-3](**3**)
	---	[6–7–1–1](**4**)	---
	---	[6–7–1–2](**3**)	---
	---	[6–7–1–3](**5**)	---
	---	[6-7-0-1]	[6-7-0-1]
	---	[6-7-0-2](**3**)	[6-7-0-2]
	---	---	[6-7-0-3](**2**)
	---	[6-7-0-4]	---

We found 7 common glycan structures between the three glycosylation sites Asn184, Asn207, and Asn241, 10 between the sites Asn184 and Asn207, 11 between sites Asn207 and Asn241, and 7 between sites Asn184 and Asn241, Table 1. The glycan structure $HexNAc_4$, Hex_6, Neu5Ac was only observed in glycosylation site Asn184; the structure $HexNAc_3$, Hex_4, Neu5Ac was only observed in the site Asn207; and the structure $HexNAc_6$, Hex_7, $Neu5Ac_3$ was observed only in the site Asn241 which matched with the findings of Zhu et al. [9] In addition, some of the common structures showed a different number of isomeric peaks among the glycosylation sites. For example, the glycan composition $HexNAc_4$, Hex_5, $Neu5Ac_2$ (4-5-0-2) was present in the three sites, but interestingly, two isomeric glycans were observed only in the site Asn207, Fig. 3. Similar results were observed for the other common glycan structures among the three glycosylation sites, Table 1. The results demonstrated the ability of our analytical strategy to unravel the glycan microheterogeneity of the glycosylation site NLFLN**207**HSE, and clearly described the distribution of all glycan structures among the three evaluated haptoglobin glycosylation sites. According to previous research, 35 of the 42 identified *N*-glycopeptides in our research were common with the reported work of Zhu et al., who evaluated similar analytical samples with an LC-EThcD-MS^2 analytical strategy and found a large number of *N*-glycopeptides; however, their work did not report isomeric structures. [9]

Principal component analysis (PCA) was performed to evaluate the ability of the obtained data to differentiate between cirrhosis and HCC. The analysis was performed with the MarkerView® (AB Sciex) software using the normalized data. The initial evaluation of the complete data showed partial separation between both cohorts, the cirrhosis and HCC samples, **Supplementary Fig. S5**. A further PCA analysis of the data by glycosylation site, Fig. 4, showed interesting results. For the sites Asn184 and Asn241, the data was not able to separate the disease cohorts as shown in Fig. 4a and Fig. 4c. In comparison, the PCA analysis for the glycosylation site Asn207 showed an important separation between cirrhosis and HCC samples. As discussed previously, the site Asn207 showed an important number of isomeric *N*-glycopeptides suggesting that the differentiation of the two cohorts is possible due to the microheterogeneous development of the glycosylation site Asn207, Fig. 4b. Further PCA comparisons between cirrhosis and early HCC (TNM stage 1) data of the site Asn207 showed important separation of the two cohorts, Fig. 4d. This results suggested that our analytical strategy is effective in differentiating between cirrhosis and early HCC. This PCA analysis used a reduced number, 7 HCC (TNM stage 1) against all cirrhosis samples, but these promising results should next be confirmed with a larger number of samples.

Heat maps were used to evaluate the changes in abundance of the individual *N*-glycopeptides quantified in cirrhosis and HCC samples. Site-specific heat maps were obtained for each glycosylation site (Asn184, Asn207, and Asn241), **Supplementary Fig. 6a - c**. The *N*-glycopeptides identified with a bold arrow indicate structures with significant changes in abundance between both diseases, cirrhosis and HCC ($p < 0.05$), with a blue line separating cirrhosis from HCC samples. The rectangles identify representative groups of glycopeptides such as sialylated and sialylated-fucosylated structures, the green color represents low abundance (down regulation), and the red color represents high abundance (up

regulation). For the site Asn184 the heat map allows us to identify two principal changes in the glycopeptide abundances between the two sample groups. The bi- and triantennary sialylated glycopeptides (MVS + 4-5-0-2, 5-6-0-1, 5-6-0-2, and 5-6-0-3) were more abundant in HCC than in cirrhosis samples. Otherwise, the bi- and triantennary sialylated glycopeptides (MVS + 3401, 4401, 6701, and 6702) were more abundant in cirrhosis than in HCC samples.

The isomeric separation achieved by our analytical strategy increased the number of glycopeptide structures in the glycosylation site Asn207 and thus the structural information in the site (microheterogeneity). The heat map for this site describes three different groups of glycopeptides with interesting changes in abundance between cirrhosis and HCC samples. The sialylated glycopeptides had different abundance patterns in this site; the triantennary mono-, di-, and trisialylated structures (NLF + 5-6-0-1, 5-6-0-2, and 5-6-0-3) were more abundant in cirrhosis than HCC samples. However, the mono-, di-, and tetra antennary mono- and disialylated glycopeptides (NLF + 3-4-0-1, 4-4-0-1, 6-7-0-1, and 6-7-0-2) were more abundant in HCC than cirrhosis samples. In the case of sialylated fucosylated glycopeptides, the heat map showed similar changes in abundance for almost all structures; these glycopeptides were low abundance in cirrhosis samples, and increased considerably for HCC samples, as expected from prior works [1,2,11]. The heat map for the glycosylation site Asn241 only showed a change in abundance in the glycopeptide structures VVL + 4-5-0-1 and 4-5-0-2. These glycopeptides were more abundant in cirrhosis than HCC samples.

The glycome changes associated with the progression of cirrhosis toward HCC was evaluated using pie graphs. For comparison, the data were separated into three main glycan groups attached to the glycopeptide sites; sialylated, sialylated fucosylated, and other structures, **Supplementary Fig. S7**. The evaluation of the complete data (sites Asn184, Asn207, and Asn241) was applied using the sum of abundances of the glycopeptides with the same glycan moiety; the changes in abundance of the identified glycan structures between the two cohorts are described in **Supplementary Fig. S7a**. Relative abundance of 91.9% and 7.4% were observed for the sialylated and sialylated fucosylated glycans respectively in cirrhosis samples. Otherwise, the relative abundance of the same glycan groups in HCC samples changed considerably to 83.4% and 16.2% respectively. The changes of the mentioned groups of glycopeptides were also evaluated by glycosylation site. As expected, according to the PCA plots and heat map analyses, the sites Asn184 and Asn241 did not show significant changes in abundance for sialylated and sialylated fucosylated glycopeptides, **Supplementary Fig. S7b** and **Fig. S7d**.

For the site Asn207, we observed relative abundance values of 87.2% and 12.1% for sialylated and sialylated fucosylated glycopeptides respectively in cirrhosis samples. In comparison, the relative abundance values of sialylated and sialylated fucosylated glycopeptides changed for HCC samples as 77.5% and 21.7%, **Supplementary Fig. S7c**. According to the results observed in the pie graphs, two main changes were identified in the glycome of serum haptoglobin from both analyzed cohorts. First is the high abundance of sialylated glycans presented in cirrhosis samples, which decreased by around 9.6% in HCC. The second important change was an increase of 8.8% in abundance of sialylated fucosylated glycans in cirrhosis samples compared to HCC samples.

The results observed in the PCA plots, heat maps, and pie graphs showed consistent changes in the haptoglobin glycopeptide heterogeneity between samples from patients with cirrhosis and HCC. The development of the microheterogeneity of the glycosylation site NLFLN207HSE of serum haptoglobin provided the position of the protein glycosylation changes to differentiate cirrhosis and HCC. Additionally, most of the changes were observed in isomeric glycopeptides, which has substantially increased the number of possible biomarker structures with respect to previous research in the field.

Biomarker Candidates as Differentiated between Cirrhosis and HCC

The statistical Wilcoxon Test was applied to the normalized data of each identified *N*-glycopeptide in samples from patients with NASH cirrhosis and NASH related-HCC, **Supplementary Table S4**. Twelve *N*-glycopeptides showed statistically significant differences in abundance between both analyzed cohorts (p – value < 0.05). To simplify the discussion, we named the following *N*-glycopeptides "biomarker candidates": two *N*-glycopeptides in the site Asn184, MVS + (4-5-1-2 and 5-6-1-3); nine *N*-glycopeptides for the site Asn207, NLF + (5-6-1-2 Iso 2, 5-6-1-3 Iso 1, 5-6-1-3 Iso 2, 5-6-0-1 Iso 2, 5-6-0-2 Iso 2, 5-6-0-3 Iso 2, 6-7-1-1 Iso 4, 6-7-1-2 Iso 1, and 6-7-1-2 Iso 3); and for the site Asn241, the *N*-glycopeptide VVL + 4-4-0-1. Eight of the biomarker candidates were sialylated fucosylated glycopeptides; among these were observed di-, tri-, and tetra-antennary structures. These indicated that the increase in abundance of sialylated fucosylated glycans in serum haptoglobin during HCC stages occurs regardless of the antennae number. Four of the biomarker candidates were sialylated glycopeptides with significant decreases in abundance from cirrhosis to HCC samples. Moreover, nine of the candidates were derived from isomeric glycopeptides, and all of them from the glycosylation site NLFLN207HSE. Figure 5 shows the Box plots of the biomarker candidates. Surprisingly, the glycopeptides NLF_6711 Isomer 4 and NLF_6712 Isomer 1 showed an unusual decrease in abundance for HCC samples. These structures presented four and three iso-forms respectively, where their decrease in abundance was caused by the ratio redistribution of the isomeric forms of these glycopeptides. The evaluation of the relative response without isomeric consideration of the structures NLF_6711 and NLF_6712 (**Supplementary Tables S2** and **S3**) demonstrated that both glycopeptides showed the expected increase in abundance in HCC samples.

AFP is a currently used biomarker in the detection of HCC, where the levels of this glycoprotein increase considerably in late stages of HCC. Unfortunately, AFP does not have sufficient sensitivity and specificity to differentiate between liver cirrhosis and **early-HCC**. To demonstrate the effectiveness of the biomarker candidates developed by our analytical strategy, receiving operating characteristic (ROC) curves were used to evaluate the sensitivity and specificity of their response, Table 2. The area under the curve (AUC) values obtained for the single biomarker candidates and AFP alone were compared, where only the glycopeptide structures NLF_5612 Isomer 2 (0.84) and NLF_5613 Isomer 1 (0.84) showed comparable AUC values with AFP (0.85). Sialylated biomarker candidates NLF_5601 Isomer 2, NLF_5602 Isomer 2, and NLF_5603 Isomer 2 were expected to show better performance based on their high abundance. However, their high standard deviation values led to a low sensitivity and specificity which is reflected

with low AUC values in the ROC analyses, Table 2 and **Supplementary Table S4**. ROC curves for the twelve *N*-glycopeptides with significant changes in abundance are shown in **Supplementary Fig. S8**.

Table 2

Area under the curve (AUC) comparing one-biomarker models and two-biomarker models (alpha-fetoprotein “AFP” + biomarker candidate) for the evaluation of cirrhosis and HCC samples “Top”. AUC for single biomarker candidates observed in the evaluation of cirrhosis and early-HCC samples “Down”. MVSHHN184LTTGATLINE = MVS, NLFLN207HSE = NLF, VVLHPN241YSQVDIGLIK = VVL, Iso = isomer, and AFP = alpha-fetoprotein. Glycan nomenclature: HexNAc, Hex, Fuc, NeuAc (*N*-acetylhexosamine, Hexose, Fucose, *N*-acetylneuraminic acid).

Single biomarker model			**Two-biomarker model**				
Marker	**AUC**	p-**value**	**AFP + Marker**	**AUC**	**CI (low)**	**CI (high)**	p-**value**
AFP	**0.85**	—	—	—	—	—	—
MVS_4512	0.73	0.331	AFP + MVS_4512	0.92	0.80	1.00	0.382
MVS_5613	0.76	0.420	AFP + MVS_5613	0.92	0.81	1.00	0.402
NLF_5612 Iso 2	0.84	0.897	AFP + NLF_5612 Iso 2	0.88	0.74	0.97	0.690
NLF_5613 Iso 1	0.84	0.930	AFP + NLF_5613 Iso 1	0.95	0.86	1.00	0.202
NLF_5613 Iso 2	0.80	0.664	AFP + NLF_5613 Iso 2	0.92	0.79	1.00	0.418
NLF_5601 Iso 2	0.74	0.346	AFP + NLF_5601 Iso 2	0.87	0.73	0.97	0.819
NLF_5602 Iso 2	0.77	0.493	AFP + NLF_5602 Iso 2	0.86	0.72	0.96	0.890
NLF_5603 Iso 2	0.72	0.282	AFP + NLF_5603 Iso 2	0.84	0.69	0.96	0.929
NLF_6711 Iso 4	0.73	0.312	AFP + NLF_6711 Iso 4	0.88	0.74	0.98	0.697
NLF_6712 Iso 1	0.72	0.296	AFP + NLF_6712 Iso 1	0.91	0.77	0.98	0.517
NLF_6712 Iso 3	0.73	0.289	AFP + NLF_6712 Iso 3	0.94	0.84	1.00	0.256
VVL_4401	0.79	0.601	AFP + VVL_4401	0.89	0.75	0.98	0.657

Single biomarker model, evaluation of cirrhosis and early-HCC samples			
Marker	**AUC**	**CI (low)**	**CI (high)**
AFP	**0.66**	0.37	0.96
MVS_5613	0.86	0.66	1.00
NLF_5612 Iso 2	0.84	0.60	1.00
NLF_5613 Iso 1	0.88	0.69	1.00
NLF_6712 Iso 3	0.80	0.55	1.00
VVL_4401	0.84	0.61	1.00

AUC values were obtained from bootstrap method (n = 1000), 95% confidence interval (CI).

To improve the performance of AFP to differentiate between cirrhosis and HCC stages, researchers have explored two-biomarker models. A Bootstrap test was applied to the response of each candidate structure + AFP to calculate the combined AUC values of the two-biomarkers. The results showed an increase in sensitivity and specificity for all combinations “biomarker candidate + AFP,” so that all combinations presented AUC values higher than AFP alone (0.85) and higher than any of the single candidates (0.84), Table 2. Six combinations showed the most significant increase in the AUC values with respect to AFP alone: AFP + NLF_5613 Isomer 1 (0.95), AFP + NLF_6712 Isomer 3 (0.94), AFP + NLF_5613 Isomer 2 (0.92), AFP + MVS_4512 (0.92), MVS_5613 (0.92), and NLF_6712 Isomer 1 (0.91). The correlation of all biomarker candidates was evaluated using the pairwise Pearson’s correlation (**Supplementary Table S5**), but even combinations of two candidates with coefficients of 0.92 and 0.89 (NLF_5602 Isomer 2 + NLF_5603 Isomer 2 and NLF_5612 Isomer 2 + NLF_5613 Isomer 1 respectively) showed considerably lower AUC values than all “biomarker candidate + AFP” curves. Regarding the sialylated biomarker candidates, it is noteworthy that these structures cannot be combined with AFP; while sialylated glycopeptides are high abundance in cirrhosis samples AFP is not, and vice versa. Therefore, these correlation differences along with the high standard deviations among samples make the sialylated glycopeptides non-viable biomarker candidates.

The sample heterogeneity of the HCC cohort allowed us to evaluated the data considering a subdivision based on **early-HCC** TNM stage 1, and seven samples for each cohort were used to calculate all statistical tests, Table 2 and **Supplementary Table S5**. Although this is a small sample number, such outstanding results must be validated. Five *N*-glycopeptides showed statistically significant differences in abundance between both cohorts, and all of these were observed in the evaluation of the whole data: one *N*-glycopeptide in the site Asn184, MVS + 5-6-1-3; three *N*-glycopeptides in the site Asn207, NLF + (5-6-1-2 Iso 2, 5-6-1-3 Iso 1, and 6-7-1-2 Iso 3); and one *N*-glycopeptide, VVL + 4-4-0-1, for the site Asn241. The area under the curve (AUC) values obtained for the single biomarker candidates and AFP were compared, Table 2. As was expected, the AUC value for AFP decreased considerably to 0.66. Without the high values reported in TNM stages 3 and 4, the selectivity and specificity of this glycoprotein decreased its effectiveness to detect HCC in early stages, Table 2. Conversely, the glycopeptides with significant changes in abundance between the two cohorts “biomarker candidates” showed substantially larger AUC values than AFP alone: the glycopeptide MVS_56123 (0.86), the glycopeptide NLF_5612 Isomer 2 (0.84), the glycopeptide NLF_5613 Isomer 1 (0.88), the glycopeptide NLF_6712 Isomer 3 (0.80), and the glycopeptide VVL_4401 (0.84), Table 2. Due to the low AUC value of AFP, the two-biomarker model “glycopeptide candidate + AFP” did not show significant increase in the AUC values of the combinations tested; rather, the AUC values observed in the ROC curves basically represented the signal of the glycopeptide candidates.

Methods

Purification of Haptoglobin from Human Serum

Serum samples were provided by the University of Michigan and the University of Texas Southwestern Hospital. These samples were archived and de-identified and were originally collected with patient agreement under IRB approvals. Informed consent was obtain from all subjects. Collection of samples was performed in accordance with the guidelines and approval of the IRB committees at both the University of Michigan and the University of Texas Southwestern Hospital. Samples represented 15 cases of liver NASH cirrhosis, and 15 cases of NASH related-HCC. The clinical information associated with the samples used in this study is summarized in Table 3. Haptoglobin was purified from a 20 µL aliquot of serum for each patient using an HPLC-based antibody-immobilized method developed in house as previously reported [1,9]. Before enzymatic digestion, the purity of the eluted haptoglobin was confirmed by 1D SDS-PAGE followed by silver staining using the ProteoSilver™ Plus Silver Stain Kit (Sigma).

Table 3
Patient clinical information.

Disease diagnosis	**Cirrhosis**	**HCC**
Number	15[a]	15[b]
Etiology	NASH	NASH
Gender % (M/F)	3 / 12	7 / 8
Age (mean ± SD)	60 ± 9	63 ± 9
AFP level[c] (median), ng/mL	2.9	6.0
ALT level[c] (mean ± SD)	38 ± 20	47 ± 27
AST level[c] (mean ± SD)	44 ± 22	71 ± 50
MELD[d] score (mean ± SD)	8 ± 1.6	12.4 ± 6.7
TNM stage, % (I,II,III,IV)	NA	47[e] / 0 / 40 / 13
[a] **Cirrhosis samples: NASH58, 60, 61, 62, 63, 64, 65, 66, 68, 69, 70, 71, 73, 74, 75.**		
[b] **HCC samples: NASH80[e], 81, 82, 83, 84, 85, 86, 87 [e]. 88 [e], 89 [e], 91, 92, 94 [e], 95 [e], 96 [e].**		
[c] **AFP, ALT, and AST levels were provided by Division of Gastroenterology and Hepatology, University of Michigan and Division of Digestive and Liver Diseases, University of Texas Southwestern.**		
[d] **MELD: model for end-stage liver disease.**		

Tryptic and Glu-C Haptoglobin Digestion

The purified serum haptoglobin was resuspended with 20 µL of 50 mM NH_4HCO_3 and denatured at 90°C in a water bath for 15 min. The denatured glycoprotein was reduced by the addition of 0.5 µL of 200 mM dithiothreitol (DTT) and incubation at 60°C for 45 min. Then, the glycoprotein was alkylated by the

addition of 2.0 µL of 200 mM iodoacetamide (IAA) and incubation at 37°C for 45 min. A second addition of 0.5 µL of 200 mM DTT and incubation at 37°C for 30 min was performed to quench the IAA excess. Trypsin was added to the treated glycoprotein sample at a concentration ratio of 1:25 and incubated at 37°C overnight. After incubation, the digestion was quenched at 90°C for 15 min. In a second enzymatic digestion, Glu-C was added to the tryptic digested sample at a concentration ratio of 1:25 and incubated at 37°C overnight. The sample was finally dried down in a SpeedVac concentrator.

LC-PRM-MS Analysis

The dried tryptic and Glu-C digests were resuspended in a solution of 2% acetonitrile (MeCN), 0.1% formic acid (FA). Five microliters of the reconstituted sample were injected onto a C18 trap column (75 µm x 10 cm, 2 µm, 100 Å; Thermo Scientific, Pittsburgh, PA) for 10 min, and the samples were then transferred to an Aclaim PepMap C18 capillary column (75 µm x 50 cm, 2 µm, 100 Å; Thermo Scientific, Pittsburgh, PA) using an Ultimate 3000 nanoUHPLC system (Dionex, Sunnyvale, CA). The flow rate was set to 300 nL/min with a temperature of 60°C. Mobile phase A was an aqueous mixture with 2% of MeCN, and 0.1% of FA, while mobile phase B was a mixture of MeCN with 0.1% of FA. The analytical gradient was 100 min long, and started at 2% mobile phase B for the initial 10 min, then increased to 38% at 11 min. During the next 70 min the organic phase B gradually developed to 60%. Subsequently, it ramped up to 90% in a period of 3 min and was maintained for 4 min. Finally, the percentage of organic phase B dropped to 2% in 1 min and was kept at that condition to pre-equilibrate the system. The nanoUHPLC system was interfaced to a Q Exactive (Thermo Scientific, San Jose, CA), and operated in positive ion mode for PRM-MS analysis. The fragmentation pattern of the precursor ions was evaluated with different collision energy (CE) levels and established as 25 eV. Then, a MS^1 scan range of 300–2000 *m/z* and a MS^2 scan range of 300–3000 *m/z* were applied to a pooled sample for the identification of the precursor ions, their fragments and retention times. All the identified precursor ions were included in the PRM-MS method with a retention time window of ± 3 min, and a mass range of ± 2 Da relative to the target mass. In order to compare the data from two cohorts, the Wilcoxon Test was performed and p values < 0.05 were considered significant. We obtained area under receiver operating characteristic curve (AUC) values, and heat maps using the SPSS® (IBM) software. PCA plots were obtained using the software MarkerView® (AB Sciex).

Conclusion

In this study, we demonstrated that site-specific *N*-glycosylation changes of serum haptoglobin can be used to differentiated NASH cirrhosis and NASH related-HCC. This assay incorporated a target parallel reaction monitoring (PRM) approach with a long C18 column capable of resolving important isomeric glycopeptide structures at high temperature. We were able to accurately quantify 72 isomeric structures corresponding to 42 glycopeptide isoforms, structures that were distributed among the haptoglobin glycosylation sites as follows: thirteen in the site Asn184, forty-four in the site Asn207, and sixteen in the site Asn241. The glycosylation site NLFLN207HSE (Asn207) showed the largest number of glycan structures, most of them with isomeric forms. The microheterogeneity of this site was strongly related

with to its ability to differentiate between the studied diseases, as can be seen in the PCA and heat map analysis. Additionally, quantitative analysis revealed twelve *N*-glycopeptides with significant changes in abundance during the progression from cirrhosis to HCC ($p < 0.05$), nine of which were located in the site Asn207, two in the site Asn184, and one in the site Asn241. Eight of the significant glycopeptides were sialylated fucosylated structures with 1 to 3 antennae, and four were sialylated structures. AUC values of ROC curves were used to compare the sensitivity and specificity of the significant glycopeptides developed by our analytical strategy against AFP. In a two-biomarker model, six combinations of "glycopeptide + AFP" showed higher AUC values than AFP alone. Further subgroup analysis of early HCC data (TNM stage 1) showed four glycopeptides with important differences with respect to AFP alone.

This study confirmed that site-specific glycoproteomic analysis is an important tool to evaluate serum haptoglobin changes during the progression of cirrhosis to HCC. Additionally, we showed that the unraveling of the haptoglobin isomeric glycopeptides contributed to enhanced the differentiation between both diseases. Moreover, by applying our LC-PRM-MS strategy we were able to identify important low abundant haptoglobin *N*-glycopeptides that are possible biomarkers to differentiate between patients with NASH cirrhosis and NASH related-HCC, and, more importantly, to differentiate between patients with cirrhosis and early stage HCC. Further validation of these markers in larger cohorts of patients is warranted to improve the early detection of HCC.

Declarations

Data availability

The mass spectrometry data generated during the current study are deposited and available in the PRIDE archive (http://www.ebi.ac.uk/pride/archive) with the dataset identifier; PXD023719.

Acknowledgement

This work was supported from the National Institutes of Health, NIH (1R01GM112490-07 (YM), 1R01GM130091-02 (YM), and 1U01CA225753-01 (DML & YM)), 1R01CA160254 (DML), R50CA221808 (JZ), and U01CA230694 (AGS). DML acknowledges support under the Maud T. Lane Professorship, JZ acknowledges the support from the National Cancer Institute.

Authors' contributions

Y.M. and D.M.L. conceived, designed, coordinated and, revised the overall study. C.D.G.R. and Y.H. carried out experiments, data analysis and drafted the manuscript. J.Zhu participated in research coordination, data analysis and revision of the manuscript. M.A. data analysis. J.Zhang performed haptoglobin purification from serum samples. S.L. and J.D. statistical analysis. N.D.P. and A.G.S. provided serum samples, clinical data and contributed to manuscript revision. The authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Huang, Y., Zhou, S., Zhu, J., Lubman, D. M. & Mechref, Y. LC-MS/MS Isomeric Profiling of Permethylated N-Glycans Derived from Serum Haptoglobin of Hepatocellular Carcinoma (HCC) and Cirrhotic Patients. *Electrophoresis* **38 (17)**, 2160–2167 (2017).
2. Zhu, J., Warner, E., Parikh, N. D. & Lubman, D. M. Glycoproteomic markers of hepatocellular carcinoma-mass spectrometry based approaches. *Mass Spectrom. Rev.* **38**, 265–290 (2019).
3. Kim, K. H. *et al.* Parallel reaction monitoring with multiplex immunoprecipitation of N-glycoproteins in human serum for detection of hepatocellular carcinoma. *Anal. Bioanal. Chem.* **411**, 3009–3019 (2019).
4. Kim, D. *et al.* Changing Trends in Etiology-Based and Ethnicity-Based Annual Mortality Rates of Cirrhosis and Hepatocellular Carcinoma in the United States. *Hepatology* **69**, 1064–1074 (2019).
5. Singal, A. G., Pillai, A. & Tiro, J. Early Detection, Curative Treatment, and Survival Rates for Hepatocellular Carcinoma Surveillance in Patients with Cirrhosis: A Meta-analysis. *PLOS Med.* **11**, e1001624 (2014).
6. Cartier, V. & Aubé, C. Diagnosis of hepatocellular carcinoma. *Diagn. Interv. Imaging* **95**, 709–719 (2014).
7. Tzartzeva, K. *et al.* Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. *Gastroenterology* **154**, 1706-1718.e1 (2018).
8. Younossi, Z. *et al.* Nonalcoholic Steatohepatitis Is the Fastest Growing Cause of Hepatocellular Carcinoma in Liver Transplant Candidates. *Clin. Gastroenterol. Hepatol.* **17**, 748-755.e3 (2019).
9. Zhu, J. *et al.* Differential Quantitative Determination of Site-Specific Intact N-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-EThcD-MS/MS. *J. Proteome Res.* **18**, 359–371 (2019).
10. Comunale, M. A. *et al.* Identification and Development of Fucosylated Glycoproteins as Biomarkers of Primary Hepatocellular Carcinoma. *J. Proteome Res.* **8**, 595–602 (2009).
11. Pompach, P. *et al.* Protein and site specificity of fucosylation in liver-secreted glycoproteins. *J. Proteome Res.* **13**, 5561–5569 (2014).
12. Gutierrez Reyes, C. D., Jiang, P., Donohoo, K., Atashi, M. & Mechref, Y. S. Glycomics and glycoproteomics: Approaches to address isomeric separation of glycans and glycopeptides. *J. Sep. Sci.* **44**, 403–425 (2020).
13. Jin, W. *et al.* Glycoqueuing: Isomer-Specific Quantification for Sialylation-Focused Glycomics. *Anal. Chem.* **91**, 10492–10500 (2019).
14. Yu, Q. *et al.* Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom.* **28**, 1751–1764 (2017).

15. Yin, H., Zhu, J., Wang, M., Yao, Z.-P. & Lubman, D. M. Quantitative Analysis of α-1-Antitrypsin Glycosylation Isoforms in HCC Patients Using LC-HCD-PRM-MS. *Anal. Chem.* **92**, 8201–8208 (2020).
16. Yin, H. *et al.* The analysis of alpha-1-antitrypsin glycosylation with direct LC-MS/MS. *Electrophoresis* **39**, 2351–2361 (2018).
17. Kurogochi, M. *et al.* Sialic Acid-focused Quantitative Mouse Serum Glycoproteomics by Multiple Reaction Monitoring Assay. *Mol. Cell. Proteomics* **9**, 2354 (2010).
18. Choo, M. S., Wan, C., Rudd, P. M. & Nguyen-Khuong, T. GlycopeptideGraphMS: Improved Glycopeptide Detection and Identification by Exploiting Graph Theoretical Patterns in Mass and Retention Time. *Anal. Chem.* **91**, 7236–7244 (2019).
19. Xiao, H., Hwang, J. E. & Wu, R. Mass spectrometric analysis of the N-glycoproteome in statin-treated liver cells with two lectin-independent chemical enrichment methods. *Int. J. Mass Spectrom.* **429**, 66–75 (2018).
20. Mysling, S., Palmisano, G., Højrup, P. & Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem.* **82**, 5598–5609 (2010).
21. Ronsein, G. E. *et al.* Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteom.* **113**, 388–399 (2015).
22. Abbatiello, S. E. *et al.* Design, Implementation and Multisite Evaluation of a System Suitability Protocol for the Quantitative Assessment of Instrument Performance in Liquid Chromatography-Multiple Reaction Monitoring-MS (LC-MRM-MS). *Mol. Cell. Proteomics* **12**, 2623 (2013).
23. Yuan, W., Benicky, J., Wei, R., Goldman, R. & Sanda, M. Quantitative Analysis of Sex-Hormone-Binding Globulin Glycosylation in Liver Diseases by Liquid Chromatography-Mass Spectrometry Parallel Reaction Monitoring. *J. Proteome Res.* **17**, 2755–2766 (2018).
24. Reiding, K. R., Bondt, A., Franc, V. & Heck, A. J. R. The benefits of hybrid fragmentation methods for glycoproteomics. *Trends Analyt. Chem.* **108**, 260–268 (2018).
25. West, C. A. *et al.* N-Linked Glycan Branching and Fucosylation Are Increased Directly in Hcc Tissue As Determined through in Situ Glycan Imaging. *J. Proteome Res.* **17**, 3454–3462 (2018).
26. Ji, E. S. *et al.* Isomer separation of sialylated O- and N-linked glycopeptides using reversed-phase LC–MS/MS at high temperature. *J. Chromatogr. B* **1110–1111**, 101–107 (2019).

Figures

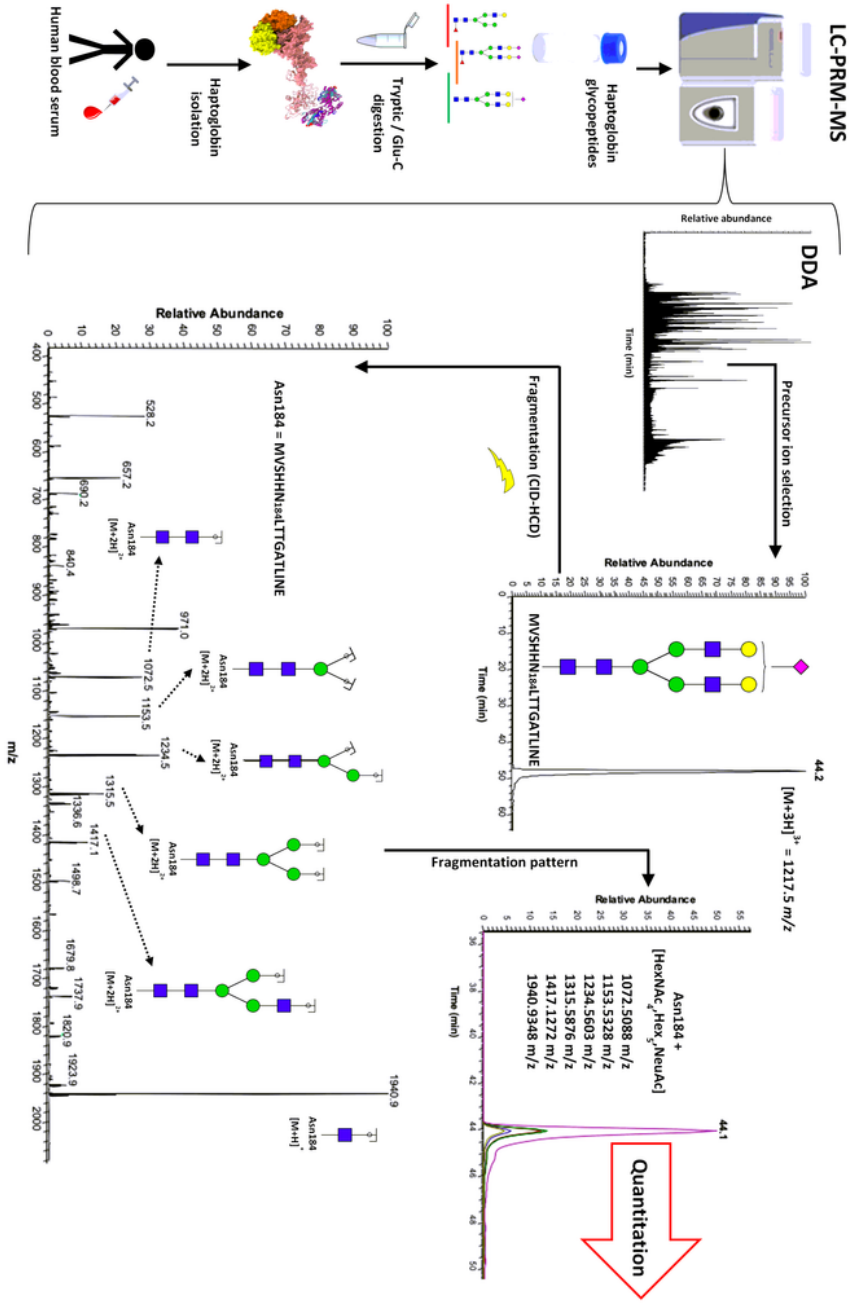

Figure 1

Workflow for the parallel reaction monitoring – mass spectrometry (PRM-MS) analysis of serum haptoglobin N-glycopeptides.

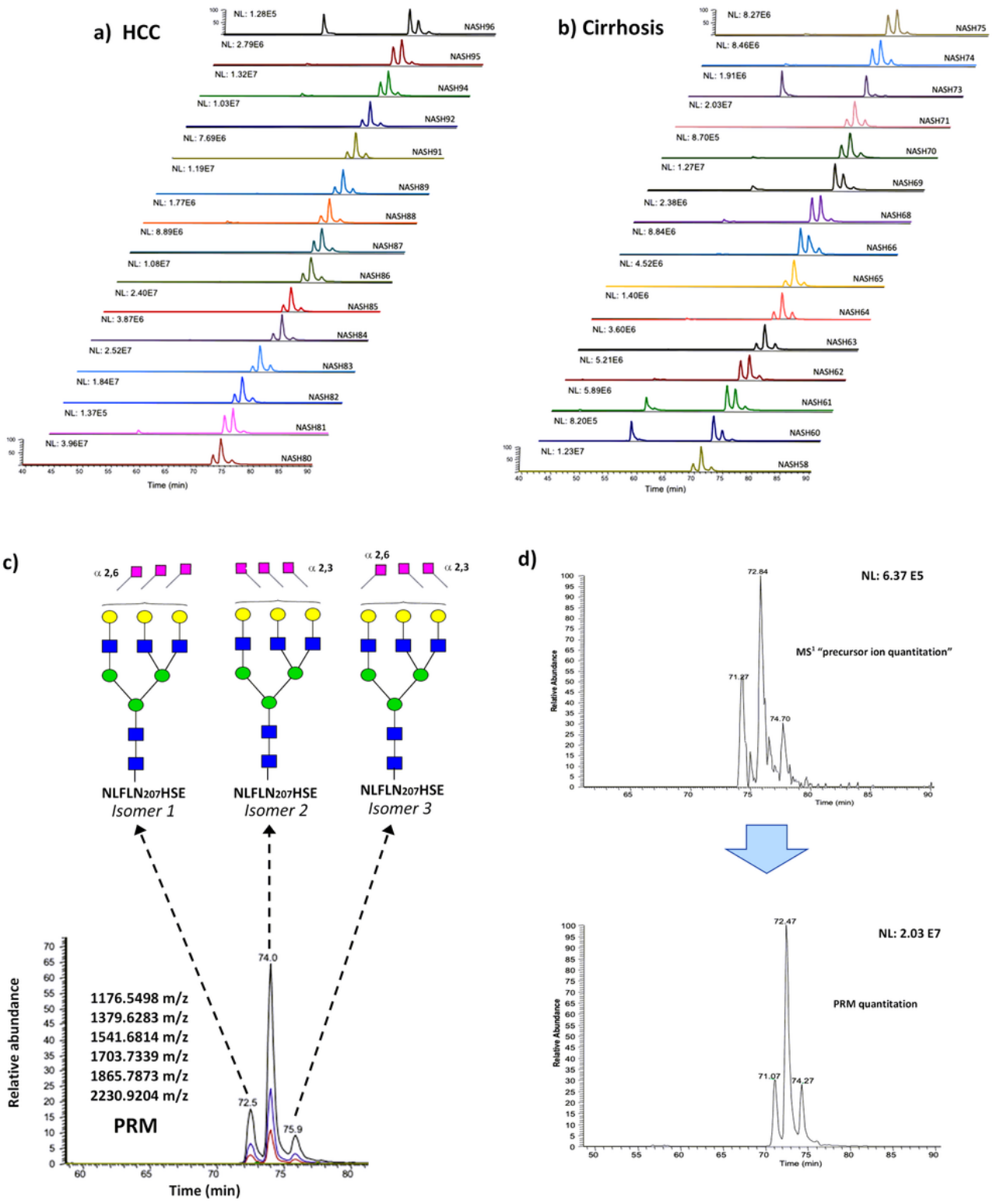

Figure 2

EICs for the glycopeptide structure NLFLN207HSE + HexNAc5,Hex6,Neu5Ac3 derived from a) cirrhosis and b) hepatocellular carcinoma patients. c) PRM quantitation of the glycopeptide structure and sialic acid linkage of the isomeric moieties. d) EICs for the glycopeptide structure NLFLN207HSE + HexNAc5,Hex6,Neu5Ac3, 1278.8295 m/z. Sensitivity difference between A) MS1 ion, and B) PRM approaches.

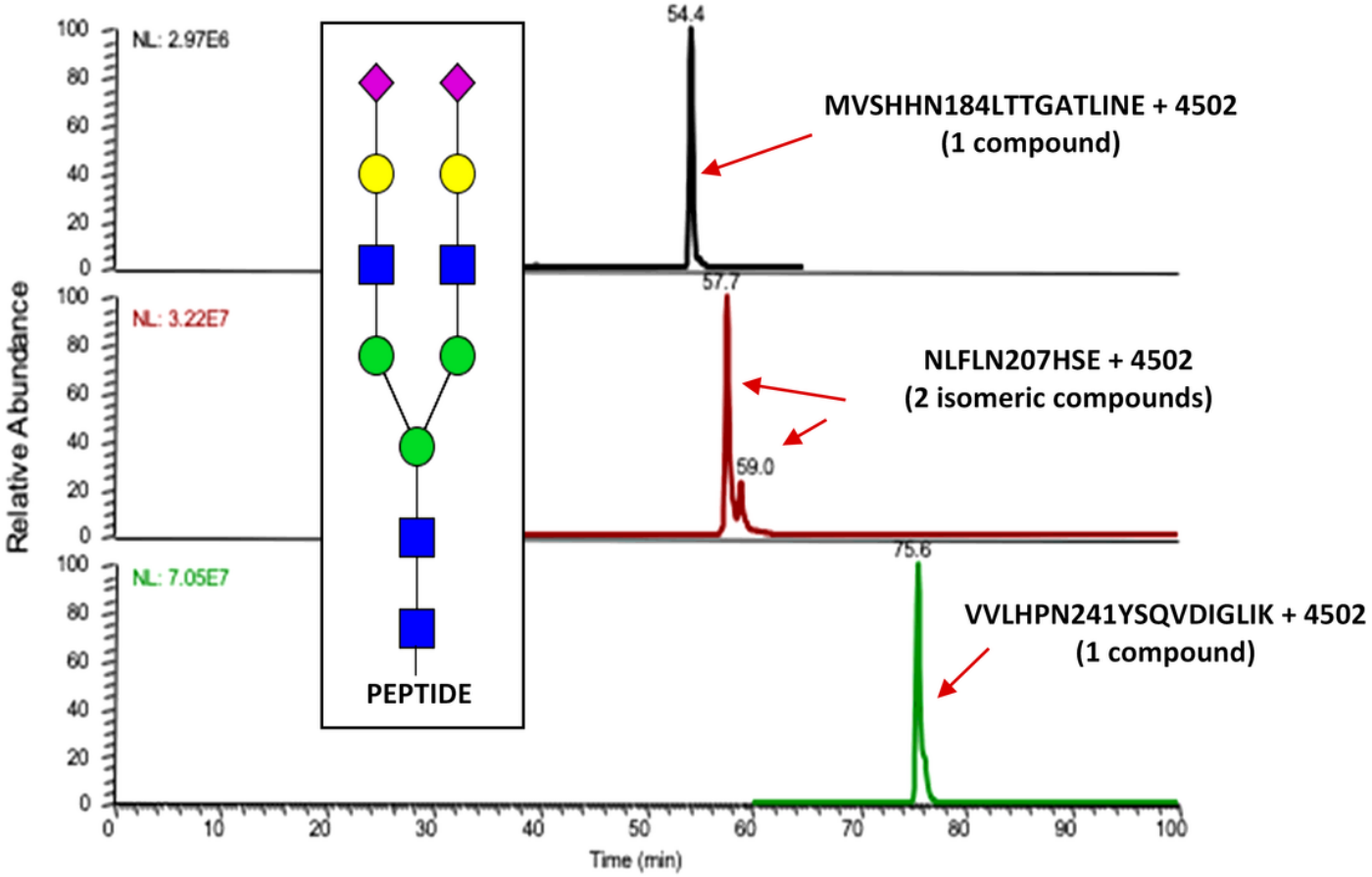

Figure 3

N-glycan distribution among glycosylation sites. Glycan nomenclature: HexNAc, Hex, Fuc, NeuAc (N-acetylhexosamine, Hexose, Fucose, N-acetylneuraminic acid).

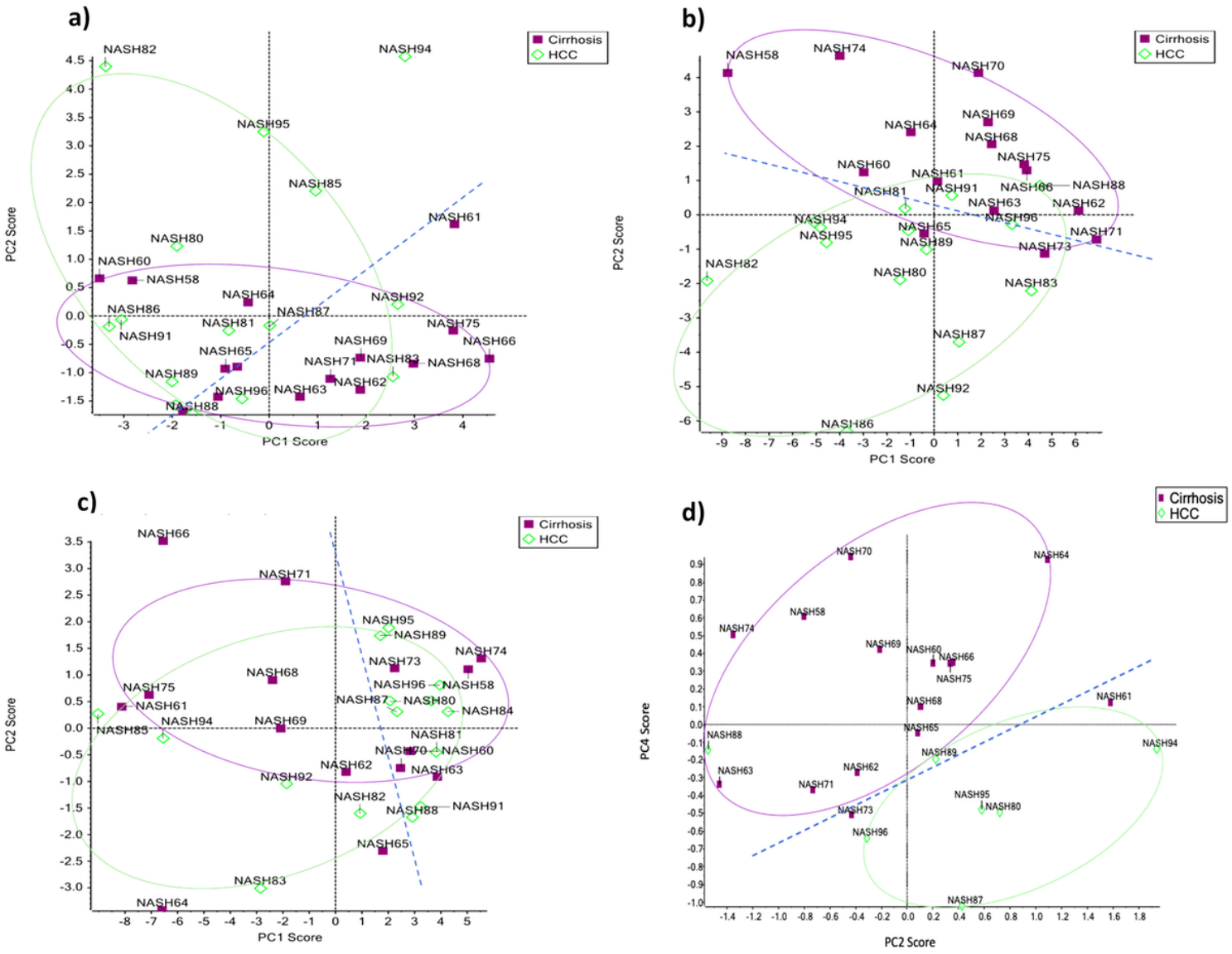

Figure 4

PCA plots for the glycosylation sites derived from human serum haptoglobin from patients with cirrhosis and HCC. a) MVSHHN184LTTGATLINE (scores for PC1 = 43.2%, and PC2 = 22.3%, autoscale), b) NLFLN207HSE (scores for PC1 = 40.5%, and PC2 = 14.7%, pareto), c) VVLHPN241YSQVDIGLIK (scores for PC1 = 69.4%, and PC2 = 9.1%, pareto), and d) NLFLN207HSE (scores for PC2 = 21.4%, and PC4 = 7.0%, Sqrt).

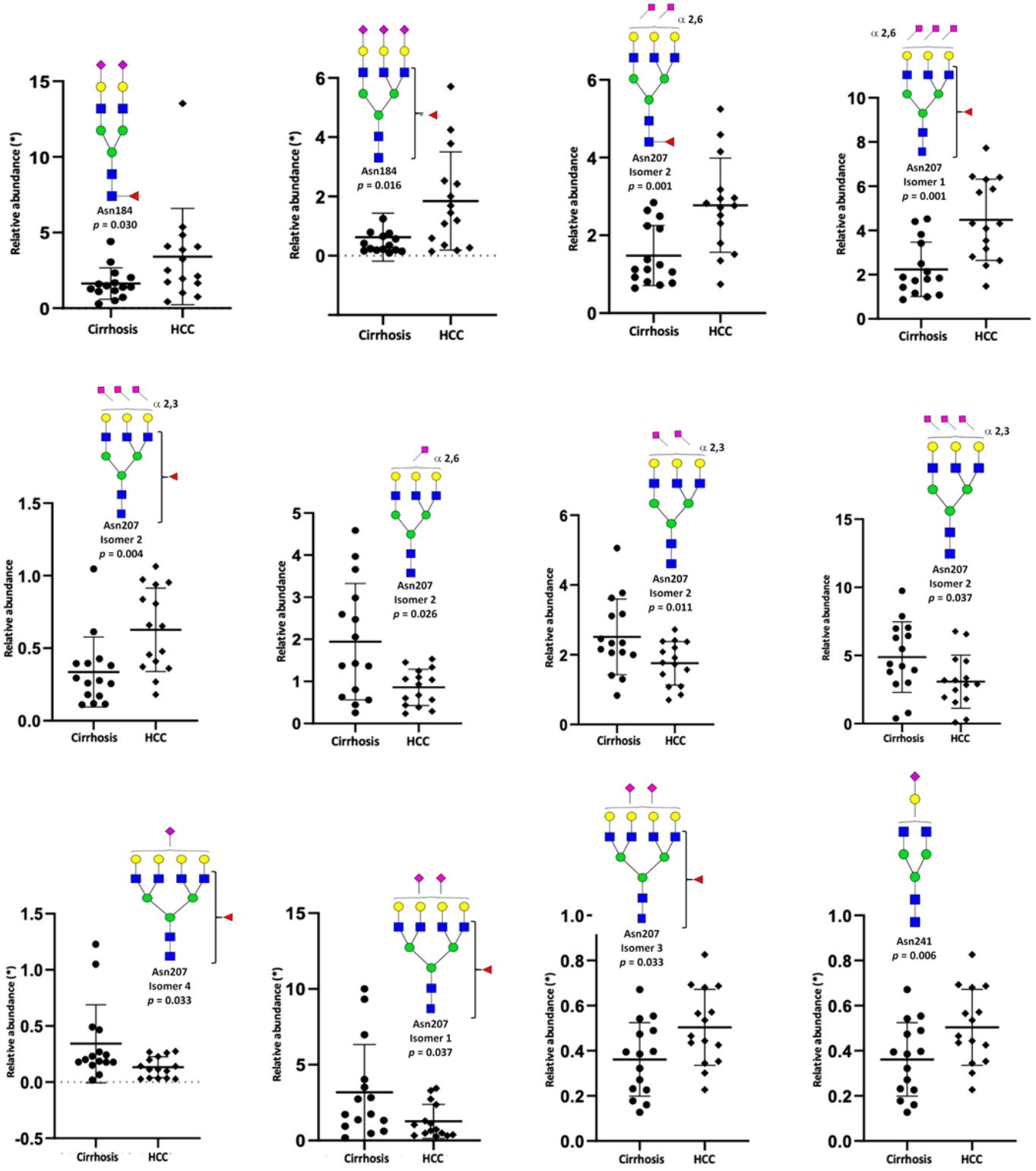

Figure 5

Glycopeptide structures with statistically significant difference between cirrhosis and HCC samples. MVSHHN184LTTGATLINE (Asn184), NLFLN207HSE (Asn207), and VVLHPN241YSQVDIGLIK (Asn241), and (*) relative abundance multiplied by 10.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PRMHaptoglobinSuplementarydataFinalVersion.docx