



# Precision N-glycoproteomics reveals elevated LacdiNAc as a novel signature of intrahepatic cholangiocarcinoma

Jun Li<sup>1</sup> , Ting Zhao<sup>1</sup>, Jing Li<sup>1</sup>, Jiechen Shen<sup>1</sup>, Li Jia<sup>1</sup>, Bojing Zhu<sup>1</sup>, Liuyi Dang<sup>1</sup>, Chen Ma<sup>1</sup>, Didi Liu<sup>1</sup>, Fan Mu<sup>2</sup>, Liangshuo Hu<sup>2</sup> and Shisheng Sun<sup>1</sup> 

<sup>1</sup> College of Life Science, Northwest University, Xi'an, China

<sup>2</sup> Department of Hepatobiliary Surgery, Institute of Advanced Surgical Technology and Engineering, The First Affiliated Hospital of Xi'an Jiaotong University, China

## Keywords

glycopeptides; glycoproteins; glycoproteomics; hepatocellular carcinoma; intrahepatic cholangiocarcinoma; mass spectrometry

## Correspondence

S. Sun, College of Life Science, Northwest University, Xi'an 710069, China

E-mail: suns@nwu.edu.cn

and

L. Hu, Department of Hepatobiliary Surgery and Institute of Advanced Surgical Technology and Engineering, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

E-mail: huliangshuo1983@hotmail.com

Jun Li and Ting Zhao contributed equally to this article

(Received 27 June 2021, revised 2 November 2021, accepted 30 November 2021, available online 18 December 2021)

doi:10.1002/1878-0261.13147

Primary liver cancer, mainly comprising hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), remains a major global health problem. Although ICC is clinically different from HCC, their molecular differences are still largely unclear. In this study, precision N-glycoproteomic analysis was performed on both ICC and HCC tumors as well as paracancer tissues to investigate their aberrant site-specific N-glycosylation. By using our newly developed glycoproteomic methods and novel algorithm, termed 'StrucGP', a total of 486 N-glycan structures attached on 1235 glycosites were identified from 894 glycoproteins in ICC and HCC tumors. Notably, glycans with uncommon LacdiNAc (GalNAc $\beta$ 1-4GlcNAc) structures were distinguished from their isomeric glycans. In addition to several bi-antennary and/or bisecting glycans that were commonly elevated in ICC and HCC, a number of LacdiNAc-containing, tri-antennary, and core-fucosylated glycans were uniquely increased in ICC. More interestingly, almost all LacdiNAc-containing N-glycopeptides were enhanced in ICC tumor but not in HCC tumor, and this phenomenon was further confirmed by lectin histochemistry and the high expression of  $\beta$ 1-4 GalNAc transferases in ICC at both mRNA and protein expression levels. The novel N-glycan alterations uniquely detected in ICC provide a valuable resource for future studies regarding to the discovery of ICC diagnostic biomarkers, therapeutic targets, and mechanism investigations.

## 1. Introduction

Primary liver cancer continues to be a growing health problem worldwide. Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) are two

major subtypes of primary liver cancer, accounting for approximately 75–85% and 10–15%, respectively [1,2]. Both together, they have been the sixth most commonly diagnosed cancer and third leading cause of cancer death globally [2]. Although being a relatively

## Abbreviations

ACN, acetonitrile; AFP, alpha fetoprotein; B4GALNT3,  $\beta$ 1-4-acetylgalactosamine transferases 3; B4GALNT4,  $\beta$ 1-4-acetylgalactosamine transferases 4; FFPE, formalin-fixed paraffin-embedded; GO, Gene Ontology; HCC, hepatocellular carcinoma; HCCP, paracancer of hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; ICCP, paracancer of intrahepatic cholangiocarcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; MALDI, matrix-assisted laser desorption ionization; MAX, mixed anion-exchange; MS, mass spectrometry; PSMs, peptide-spectrum matches; RT, room temperature; TCGA, The Cancer Genome Atlas; TFA, trifluoroacetic acid; TMT, tandem mass tag; WFA, wisteria floribunda agglutinin.

rare cancer, ICC has caused more and more attention with increasing incidence and mortality worldwide [3].

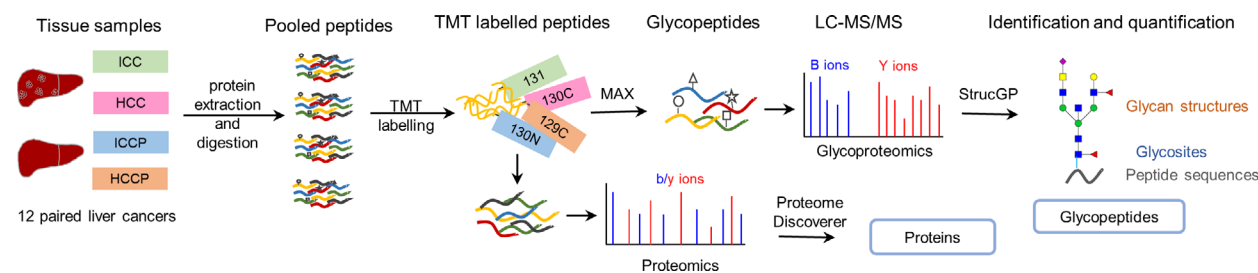
Intrahepatic cholangiocarcinoma arises from the intrahepatic biliary tract and is different from HCC with respect to morphology, metastatic capacity, and their responses to cancer therapy [4]. Distinguishing between ICC and HCC has been a major clinical issue because their management and prognosis differ significantly; however, the molecular pathogenesis underlying these differences is largely unclear [5]. In contrast to the well-characterized HCC, substantially less is known about ICC. In fact, many ICC patients are actually detected during HCC screening or treatment such as hepatic resection or transplantation for a presumed diagnosis of HCC and such misdiagnosis can then lead to inappropriate use of treatment aimed at HCC [5,6]. Although more research have been recently reported in terms of genetic, epigenetic, chromosomal, and microRNAs aberrations in ICC [7], the understanding of the molecular abnormalities involved in ICC pathogenesis is still limited. Exploring the molecular alterations underlying the carcinogenesis of ICC as well as the molecular differences between ICC and HCC will be beneficial for the differential diagnosis and targeted clinical treatment of ICC patients.

Glycosylation is one of the most important and complex post-translational modifications of proteins [8,9]. It is involved in diverse biological processes, such as recognition and regulatory functions, cell communication, adhesion, immune defense, as well as cell growth and development. Abnormal glycosylation has been recognized as an essential feature of many cancers and plays a fundamental role in key pathological steps of tumor development and progression [10]. In HCC, a number of aberrant glycosylation have been reported [11]. For example, the increased level of  $\alpha$  2,6-linked sialylation and tetra-antennary N-glycans is closely related to HCC progression [12,13]. In addition, the core-fucosylated form of alpha fetoprotein (AFP-L3) has been used as an important serum biomarker for HCC diagnosis [14]. However, only limited lectin-specific glycans (sWGA-specific glycans) or glycoproteins (WFA-Mucin 1 and WFA-LICAM) have been detected in ICC without detailed glycan structure and glycoprotein information [15–17]. A systematic analysis is necessary to explore the glycosylation changes in ICC and to discriminate ICC from HCC tumors.

During the past few decades, mass spectrometry-based glycomic approaches have been widely used for the large-scale identification of abnormal glycosylation in cancers including HCC [11,13]. Released N-glycans in HCC tissues have been widely studied by glycomics

based on matrix-assisted laser desorption ionization imaging mass spectrometry (MALDI-IMS) [18,19], MALDI-TOF-MS [20], HPLC [13], LC-MS/MS [21,22], and DNA sequencer-assisted carbohydrate electrophoresis [23], while no MS-based glycomic analysis has been reported in ICC tissues. These glycomic methods could obtain detailed compositions and/or structure analysis of released glycans; however, they lost the glycoprotein and glycosite information. Recently, intact glycopeptide-based glycoproteomic methods and software have been developed for the simultaneous identification of glycan, corresponding glycosite and glycoprotein. It has also been applied on HCC tissues/cell lines and revealed site-specific N-glycan changes in HCC [24–26]. Unfortunately, most currently used software relies on glycan database for glycopeptide interpretation and mainly provides information on glycan compositions rather than functional structures. Recently, we have developed a novel glycan database-independent strategy to determine detailed N-glycan structures on intact glycopeptide through the combined utilization of B and Y ions in tandem mass spectra (MS/MS) with low HCD collision energy [27]. It enables the high-throughput *de novo* structural sequencing of site-specific N-glycans on intact glycopeptide using software termed STRUCGP [27], which has been applied on the HCC cell lines and revealed site-specific changes of fucosylated N-glycans during epithelial–mesenchymal transition of HCC cells [28].

In this study, we combined the newly developed N-glycoproteomic methods with stable isotopic labeling-based quantification and applied them on tumor tissues of ICC and HCC to investigate the differences in aberrant site-specific N-glycosylation between ICC and HCC tumors (Fig. 1). The paired paracancer tissues (ICCP and HCCP) were also included as controls to determine the site-specific glycosylation changes in ICC and HCC tumors. The detailed information of the 12 paired samples is summarized in Table S1. Briefly, the trypsin-digested peptides of individual tumor/paracancer tissue from the same sample group were pooled into one sample and four pooled samples (including ICC, ICCP, HCC, and HCCP groups) were labeled with the tandem mass tag (TMT) reagents [29]. Glycopeptides were then enriched from TMT-labeled peptide mixture using mixed anion-exchange columns (MAX) [30] and analyzed by LC-MS/MS and STRUCGP [27], which enables large-scale and simultaneous identification of N-glycan structures, N-glycosites, and modified N-glycoproteins. By extracting TMT reporter ion intensities from their identified MS/MS spectra, the intact N-glycopeptides were quantified among four groups. The differences of aberrant site-specific N-



**Fig. 1.** Workflow of quantitative glycoproteomic and proteomic analyses of tissue samples from intrahepatic cholangiocarcinoma (ICC) and hepatocellular carcinoma (HCC). Proteins were extracted from six pairs of tumors and paracancer tissues from ICC and HCC. Equal amounts of tryptic peptides were pooled into four sample groups and labeled with four channels of the TMT reagents. The peptides were combined and separated into two aliquots: one aliquot for intact glycopeptide enrichment and the other aliquot for direct proteomic analysis. Both intact glycopeptides and peptides were analyzed by LC-MS/MS and quantitatively identified using StrucGP and Proteome Discoverer for site-specific glycosylation and proteomic analysis, respectively.

glycans in ICC and HCC were systematically demonstrated by analyzing their glycan features as well as their modified glycoproteins. The proteomic analysis was also performed to further investigate whether those glycopeptide changes occurred at protein expression or site-specific glycosylation level. The data will enhance our understanding of glycosylation differences between ICC and HCC that might be associated with their different carcinogenesis and progression.

## 2. Materials and methods

### 2.1. Tissue samples

Tissue samples were prospectively collected from patients undergoing ICC or HCC resection at the First Affiliated Hospital of Xi'an Jiaotong University, China, from 2014 to 2016. The study was approved by Human Ethics Committee at the First Affiliated Hospital of Xi'an Jiaotong University. In total, 12 tumor samples were collected from patients with ICC ( $n = 6$ ) and HCC ( $n = 6$ ), and their paired paracancer tissues ( $n = 12$ ) were also collected as control samples. The age range of the participants was between 20 to 68 years old, with an average age of 54.8. Besides, serum AFP concentrations were negative in all ICC patients and randomly distributed from low to high in HCC patients. The detailed clinical information of samples is summarized in Table S1. All tissue samples were stored at  $-80^{\circ}\text{C}$  until use. All the study methodologies conformed to the standards set by the Declaration of Helsinki.

### 2.2. Protein extraction and digestion

Liver tissues ( $\sim 50$  mg tissue per sample) were cut into pieces and washed twice with precooled PBS to remove

serum. Then, tissues were suspended with 8 M urea/1 M  $\text{NH}_4\text{HCO}_3$  buffer, homogenized using an automatic rapid tissue homogenizer (60 Hz, Shanghai Jing Xin, China) and sonicated by ultrasonic cell distribution system until the upper solution was clear. The protein concentrations were measured by BCA reagent (Beyotime, Shanghai, China). The proteins were then reduced by 5 mM dithiothreitol (DTT) at  $37^{\circ}\text{C}$  for 1 h and alkylated by 15 mM iodoacetamide in the dark for 30 min at room temperature (RT), followed by addition of another 2.5 mM DTT for 10 min at RT. Protein pellet solutions were digested overnight by sequencing grade trypsin (Promega, Madison, WI, USA; protein: enzyme, 100 : 1) in 2 M urea/0.25 M  $\text{NH}_4\text{HCO}_3$  buffer at  $37^{\circ}\text{C}$ . The digested solutions were acidified with trifluoroacetic acid (TFA) to  $\text{pH} < 2$  and centrifuged at 15 000  $g$  to remove any particulate matter. The peptides were desalted with Oasis HLB column (Waters, Milford, MA, USA) and eluted in 1 mL solution of 60% acetonitrile (ACN)/0.1% TFA. The peptide concentrations were measured by UV absorbance at 215 nm using a DS-11 spectrophotometer (Denovix, Wilmington, DE, USA).

### 2.3. Tandem mass tag labeling

Equal amounts of tryptic peptides from each tumor or paracancer tissue within the same sample group (six samples per group) were pooled into one sample prior to TMT labeling. Each of four pooled samples (including ICC, ICCP, HCC and HCCP pooled samples) was labeled by one channel of the  $\text{NH}_2$ -reactive N-terminal labeling reagents (10-plex TMT; Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocols [TMT channel of 129C (TMT-129C): HCC paracancer; TMT-130C: ICC paracancer; TMT-130N: HCC tumor; TMT-131: ICC tumor]. The concentration of TMT-labeled peptides was measured by BCA

assay. Then, equal amounts of the labeled peptides with four channels were combined and desalted by HLB columns prior to LC-MS/MS analysis.

#### 2.4. Enrichment of N-glycopeptides using mixed anion-exchange extraction cartridges

The intact glycopeptides were enriched from TMT-labeled peptide samples using mixed anion-exchange columns (MAX, Waters) [30]. Briefly, the tryptic peptide mixtures eluted from HLB column were adjusted to a final concentration of 95% ACN/1% TFA. Peptides were loaded onto activated MAX cartridges twice and washed with 95% ACN/1% TFA for three times. Enriched glycopeptides were eluted in 400  $\mu$ L solution of 50% ACN/0.1% FA, dried by vacuum concentration with an RVC 2-18 CDplus concentrator (Christ, Osterode am Harz, Germany), and resuspended in 20  $\mu$ L of 0.1% FA solution for LC-MS/MS analysis.

#### 2.5. LC-MS/MS analysis

Each sample underwent triplicate LC-MS/MS runs on an Orbitrap Fusion Lumos Mass Spectrometer equipped with an Easy-nLC™ 1200 system (Thermo Fisher Scientific, San Jose, CA, USA). About 1  $\mu$ g labeled peptide or intact glycopeptide was separated by LC with the use of an Acclaim PepMap100 C18 column (50 cm, 75  $\mu$ m i.d., 3  $\mu$ m) protected by a pre-column (2 cm, 75  $\mu$ m i.d., 3  $\mu$ m). The flow rate was kept at 200 nL·min<sup>-1</sup> with mobile phase consisting of 0.1% FA in water (A) and 0.1% FA in 80% ACN (B).

The gradient profile (110 min) for proteomics was set as follows: 3–7% B for 1 min, 7–35% B for 84 min, 35–68% B for 15 min, 68–100% B for 1 min and equilibrated in 100 B for 9 min. The spray voltage was set at 2.4 kV. Orbitrap MS1 spectra (AGC  $4 \times 10^5$ ) were collected from 350–1800  $m/z$  at a resolution of 60K followed by data-dependent HCD MS/MS (resolution 50K, collision energy 37%) using an isolation width of 1.6 Da. Peptides with charge states from 2 to 6 were selected for MS/MS acquisition. A complete gradient profile of 120 min for glycoproteomics was set as follows: 3–7% B for 1 min, 7–35% B for 90 min, 35–68% B for 19 min, 68–100% B for 1 min and equilibrated in 100 B for 9 min. The spray voltage was set at 2.4 kV. Orbitrap MS1 spectra (AGC  $4 \times 10^5$ ) were collected from 600 to 1800  $m/z$  at a resolution of 60K followed by data-dependent HCD MS/MS (resolution 50K, collision energy 20% and 37%) of the 20 most abundant ions using an isolation width of 1.6 Da. Peptides with charge states from 3 to 7 were selected for MS/MS acquisition.

#### 2.6. Protein identification and quantification

The LC-MS/MS data of global protein were searched against UniProt human protein databases by Sequest in Proteome Discoverer (Thermo Fisher Scientific, version 2.3) using same parameters described previously [26]. Briefly, the search parameters were set as follows: up to two missed cleavages were allowed for trypsin digestion, 10 p.p.m. and 0.02 Da mass tolerances were set for precursor and MS/MS ions, respectively; carbamidomethylation (C, +57.021464 Da) and 10-plex TMT (N-termini of peptides, +229.162932 Da) were set as static modifications; oxidation (M, +15.9949 Da), N-termini acetylation (+42.010565 Da), and 10-plex TMT (K, +229.162932 Da) were set as dynamic modifications. All results were filtered with 1% false discovery rate (FDR). The peptides were quantified based on intensities of TMT reporter ions (normalized by total intensities within the TMT set) in their identified MS/MS spectra.

#### 2.7. Intact glycopeptide identification and quantification

The identification of intact glycopeptides was performed using STRUCGP [27]. Briefly, the LC-MS/MS raw data of glycopeptides were first converted to 'mzML' format using Trans-Proteome Pipeline [31] and searched against the built-in glycan branch structure database from STRUCGP [27] and UniProt human protein database with the same parameters as mentioned above for proteomic data. The potential glycosite-containing peptides were screened with the N-X-S/T motif (X is any amino acid except Proline). Other parameters were set as follows: at least two oxonium ions out of the top 10 fragment ions in the MS/MS spectra were used for extraction of MS/MS spectra for intact glycopeptide. The mass tolerances of 10 and 20 p.p.m. were allowed for precursors and fragmentation ions. The identification results were filtered with 1% FDR for both peptide sequences and glycan structures, which was estimated by the decoy peptide method [32,33] and decoy spectra method [27], respectively. The search results were ranked based on their scores, and the peptide/glycan with the highest score was considered to be the corrected identification. In addition to FDR controls at both peptide and glycan levels, a probability strategy was used to further evaluate the reliability of each module of identified glycan structures [27].

The quantification information of intact glycopeptides was extracted from their identified MS/MS spectra based on intensities of TMT reporter ions. The intensity of TMT reporter ions was first normalized

using the global scaling factors obtained from global proteomic results to adjust its intensity across the four channels and the normalized intensity was used for quantitative analysis of identified glycopeptide. Quantification data were further filtered with at least five peptide-spectrum matches (PSMs) per glycopeptide. Median intensities of all PSMs were calculated for each sample and used to calculate the ratios of intact glycopeptide between samples.

## 2.8. Gene ontology and KEGG pathway analyses

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed for differentially expressed glycoproteins in ICC and HCC tissue samples using STRING: functional protein association networks (<https://string-db.org>). The thresholds of count > 2 and *P*-value < 0.05 were applied as filters on GO and KEGG pathway analyses of altered glycoproteins.

## 2.9. Other bioinformatic analysis

Multivariate statistic principal component analysis was created in R 3.2.2 using the FACTOMINER (<http://CRAN.R-project.org/>) [34] for four groups of samples (tumor and paracancer tissues of ICC and HCC) based on their intact glycopeptide expression profiles. For two-way hierarchical clustering analysis, log<sub>2</sub> fold change of abundance ratios of differentially expressed glycopeptides among four different groups was clustered using the PHEATMAP package in R language using Euclidean as a distance measure for column and row clustering.

## 2.10. Fluorescence-based lectin histochemistry and immunohistochemistry

Formalin-fixed paraffin-embedded (FFPE) tissue sections from the paired tumor and paracancer tissue of ICC were dewaxed and hydrated with xylene and gradient alcohol, respectively. For lectin histochemistry, Cy5-labeled WFA was applied to detect the LacdiNAc structures according to previously described protocols [35] with some modifications. Briefly, the sections were washed, blocked with 5% (w/v) BSA and 0.04% Triton for 1 h and incubated with 1 μg·μL<sup>-1</sup> Cy5-labeled WFA overnight in the dark at RT, followed by staining with DAPI (1 μg·mL<sup>-1</sup> in PBS; Solarbio, Beijing, China). The fluorescence images were acquired by a fluorescence microscopy, and the relative intensity of fluorescence was quantified by IMAGEJ software (<http://imagej.net>).

For immunohistochemistry, the pretreated FFPE tissue sections were blocked and incubated with anti-

B4GALNT3 antibody (orb317451 Biorbyt, UK) or anti-B4GALNT4 antibody (orb546266; Biorbyt, Cambridge, UK) at a 1 : 100 dilution, followed by detection using the PV and DAB chromogenic kits (Servicebio, Wuhan, China). The slides were scanned using Leica SCN400 slide scanner (Leica Biosystems, Wetzlar, Germany), and five different fields of view were randomly selected in each section for quantification.

## 2.11. Gene expression analyses in public datasets

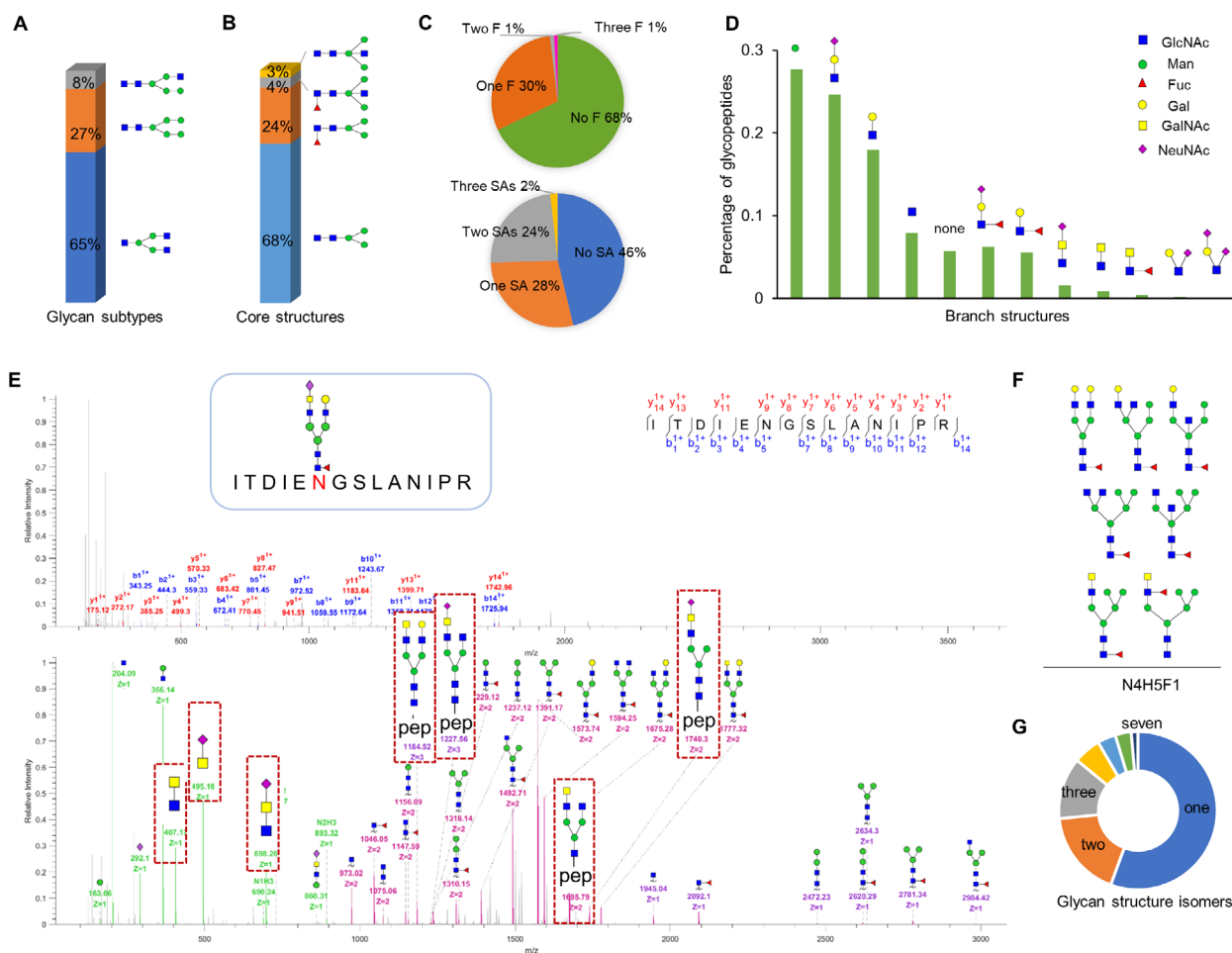
The transcriptome profiling data and clinical data were obtained from The Cancer Genome Atlas (TCGA) Data Portal (<https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>). The expression matrix of B4GALNT3 and B4GALNT4 in ICC and HCC patient tissues were obtained from the gene expression of RNAseq (HTSeq-Counts) dataset for cholangiocarcinoma (CHOL) [36] and liver hepatocellular carcinoma (LIHC) [37]. In order to exclude other cancers, samples belong to ICC and HCC were further extracted from the CHOL and LIHC datasets according to the primary pathology classification in clinical information. There are 41 samples available for ICC (33 tumors and 8 normal tissues) and 413 samples for HCC (364 tumors and 49 normal tissues). The raw data of gene expression levels were log<sub>2</sub>(*x* + 1)-transformed and processed using DESEQ2 package in R language. Statistical significance of the expression differences between groups was determined using Wilcoxon's test.

## 3. Results and Discussion

### 3.1. Structural profiling of site-specific N-glycans in ICC and HCC

By using our recently developed glycoproteomic method and glycan database-independent software STRUCGP, a number of 4741 unique intact N-glycopeptides were identified within 1% false discover rate (FDR) from TMT-labeled tumor and paracancer tissue samples of ICC and HCC (Table S2). These intact glycopeptides consist of 486 N-glycans structures (221 compositions) modified at 1235 glycosite-containing peptides from 894 N-glycoproteins.

Based on module structures of the glycans, the identified N-glycans in ICC and HCC were composed of four types of core structures and 12 branch structures with three glycan subtypes. In terms of glycan subtypes, more than 65% of glycosites were occupied by



**Fig. 2.** Large-scale profiling of intact glycopeptides in ICC and HCC samples ( $n = 24$ ). (A) Proportions of different glycans based on glycan subtypes. (B) Proportions of four types of core structures in ICC and HCC. (C) Intact glycopeptide PSMs of different glycans based on the numbers of fucose and sialic acid modifications. PSM, peptide-spectrum matches; F, fucose; SA, sialic acid. (D) Twelve branch structures identified in ICC and HCC and their relative abundances. (E) Representative MS/MS spectra for identification of LacdiNAc-containing glycopeptide: peptide ITDIE<sup>n</sup>GSLANIPR modified by a LacdiNAc-containing glycan (N5H4F1S1) from Asporin (ASPNI). <sup>n</sup> indicates the glycosylation site. The sequence of peptide was identified by matched b and y ions (labeled in blue and red, respectively) in MS/MS spectrum with high HCD energy (HCD = 37, upper spectrum). The glycan structure was identified by B and Y ions (labeled in green and purple, respectively) in MS/MS spectrum with low HCD energy (HCD = 20, bottom spectrum). (F) Seven different glycan structures discriminated from one glycan composition (N4H5F1). (G) Number of glycan structure isomers identified in ICC and HCC tissues.

complex glycans, followed by oligo-mannose glycans (27%), and hybrid glycans (Fig. 2A). Among the core structures, the typical core structure HexNAc2Hex3 accounted for 68% of all unique glycopeptides, followed by fucosylated core structure (24%) and bisecting core structure (Fig. 2B). In addition, about one-third of the glycosites (32%) were occupied by fucosylated glycans with one to three fucoses (Fig. 2C) and 92.7% of identified fucosylated glycans were assigned as core-fucosylated, which indicated a general occurrence of core fucose in ICC and HCC tissues. More than half of glycosites (53.8%) were modified by sialylated glycans containing 1–3 sialic acids (Fig. 2C). The

percentages of fucosylation and sialylation identified in this study were consistent with previously reported glycoproteomic results in HCC [26]. Among the 12 identified branch structures, the common LacNAc-containing (HexNAc + Hex) branches accounted for most of intact glycopeptides: LacNAc with or without Neu5Ac took 43%, LacNAc with or without fucose took 12% (Fig. 2D). Interestingly, the uncommon LacdiNAc-containing (HexNAc + HexNAc) branches with or without Neu5Ac/fucose were also identified by characteristic B ions containing two HexNAc (HexNAc2, HexNAc2Sia1, or HexNAc2Fu1) (Fig. 2D). The characteristic B ions have already been used for

identification of LacdiNAc-containing structures on purified glycoproteins such as prostate-specific antigen [38] and the SARS-CoV-2 S protein [39]. However, no direct identification of LacdiNAc structures has been reported in previous glycomic or glycoproteomic analysis of HCC/ICC tissues/cell lines [18–26].

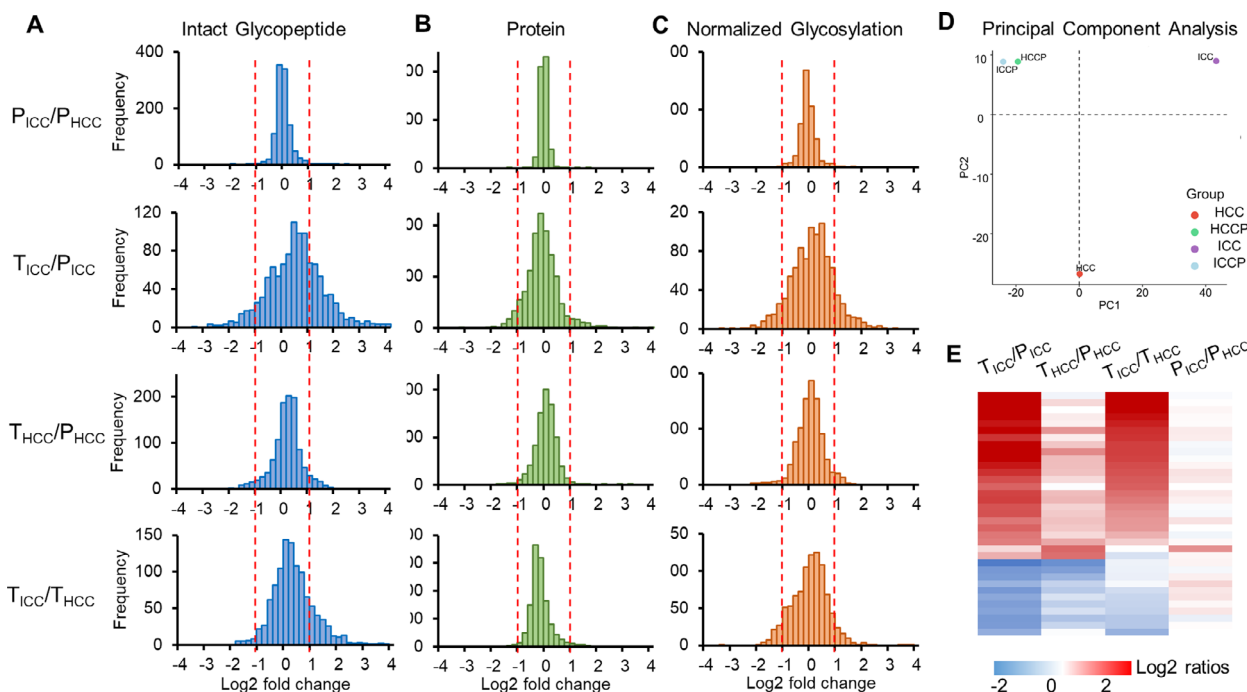
The identification of LacdiNAc-containing branches further leads to identification of LacdiNAc-containing glycopeptides by characteristic B ions and corresponding Y ions with low energy HCD fragmentation. Representative MS/MS spectra in Fig. 2E shows the identification of glycopeptide with sialylated LacdiNAc by a series of feature B ions (HexNAc2 ion at  $m/z$  407.17, HexNAc2Sia1 at  $m/z$  698.26, HexNAcSia1 at  $m/z$  495.18) and Y ions (such as peptide+HexNAc4Hex3-Fuc1Sia1 at  $m/z$  1740.3, peptide+HexNAc5Hex4Fuc1 ion at  $m/z$  1184.52, and peptide+HexNAc5Hex3Fuc1-Sia1 at  $m/z$  1227.56). Among the 486 identified N-glycans, up to seven different glycan structure isoforms can be discriminated from one single composition (Fig. 2F) and nearly half of the glycan compositions could be sorted into at least two distinct glycan isoforms (Fig. 2G). Furthermore, some glycan isoforms attached at the same peptide (but not all) could be directly

separated by the C18 column. For example, different isomers of the glycan with a composition of HexNAc5Hex4Fuc1 could be eluted at separated retention time in the LC–MS (Fig. S1). Overall, structural characterization of site-specific N-glycans has great potential to identify novel aberrant glycosylation in ICC and HCC.

### 3.2. ICC is different from HCC tumors at both glycoprotein and site-specific glycosylation levels

Based on intensities of TMT reporter ions, the intact N-glycopeptide abundances among ICC, HCC, and their paired paracancer tissues were quantified and compared. The TMT-based glycoproteomic enables the quantification of glycopeptides with isomeric N-glycans by directly extracting the abundances of TMT reporter ions from the MS2 spectra of identified glycopeptides. To improve the quantification accuracy, only glycopeptides with at least five PSMs were selected, which resulted in quantification of 1084 glycopeptides from 212 glycoproteins (Table S3).

We first compared the glycopeptide abundances between two groups of paracancer tissues, which indicated that over 98% of N-glycopeptides were within



**Fig. 3.** Relative quantification of intact glycopeptides, proteins, and glycosylation occupancies among different sample groups ( $n = 24$ ). (A–C) The frequencies of intact glycopeptides (A), proteins (B), and glycosylation occupancies (C) were compared between paracancer tissues of ICC and HCC (upper, PICC/PHCC), ICC tumors and paired paracancer tissues (TICC/PICC), HCC tumors and paired paracancer tissues (THCC/PHCC), as well as ICC and HCC tumors (TICC/THCC). (D) Principal component analysis of intact glycopeptides identified from ICC and HCC tumors as well as their paired paracancer tissues (ICCP and HCCP). (E) Heatmap of altered glycoproteins in ICC or HCC tumors compared with their paired paracancer tissues at protein expression level.

twofold changes (Fig. 3A). By using twofold change as a cutoff, we found that when compared with paired paracancer samples, 37.9% and 10.6% of N-glycopeptides changed in ICC and HCC, respectively. A direct comparison between ICC and HCC showed that 21.1% of glycopeptides changed between ICC and HCC (Fig. 3A). These results indicated that differences of glycopeptides existed not only between tumors (ICC and HCC) and their paired paracancer tissues, but also between ICC and HCC tumors. This observation was further confirmed by the principal component analysis of intact glycopeptides among four groups, which indicated that ICC tumors, HCC tumors, and their paracancer tissues are located in three distinct regions (Fig. 3D).

At the global proteome level, 10% and 2.2% of 2545 quantified proteins varied more than twofold in ICC and HCC tumors in comparison with their paired paracancer tissues, respectively (Fig. 3B). A comparison between ICC and HCC showed that 3.6% of proteins varied more than twofold. The results indicated a smaller degree of change at proteomic level. By focusing on glycoproteins identified in both glycoproteomic and proteomic data, we found that 36 glycoproteins were altered ( $\geq 2$ -fold change) in ICC and/or HCC compared with their paired paracancer tissues. By excluding any glycoprotein that changed between two groups of paracancers, a further difference ( $\geq 2$ -fold change) of 17 glycoproteins was observed between ICC and HCC tumors including 16 up-regulated and one down-regulated glycoproteins in ICC (Fig. 3E and Table S4). GO analysis on these 17 altered N-glycoproteins showed that they were mainly located in extracellular matrix region, followed by Golgi apparatus part, lysosomal lumen, and endoplasmic reticulum part (Fig. S2), which is consistent with the normal distribution of N-glycoproteins. KEGG analysis showed that they were mainly enriched in pathways including proteoglycans in cancer, TGF-beta signaling pathway, ECM-receptor interaction, and focal adhesion (Fig. S2). All those pathways are important cellular processes regulating cell proliferation, adhesion and survival and the abnormalities of them are closely connected to various types of cancers [40–42].

To determine whether these glycopeptide changes occurred at the protein or glycosylation levels, we calculated normalized glycosylation by dividing intact glycopeptide ratios with their glycoprotein ratios. Using the approach, we found that the normalized glycosylation on 20.1% and 6.2% of glycopeptides changed more than twofold in ICC and HCC tumors compared with their paired paracancer tissues, respectively (Fig. 3C). The comparison between ICC and HCC tumor samples showed that 13.1% of

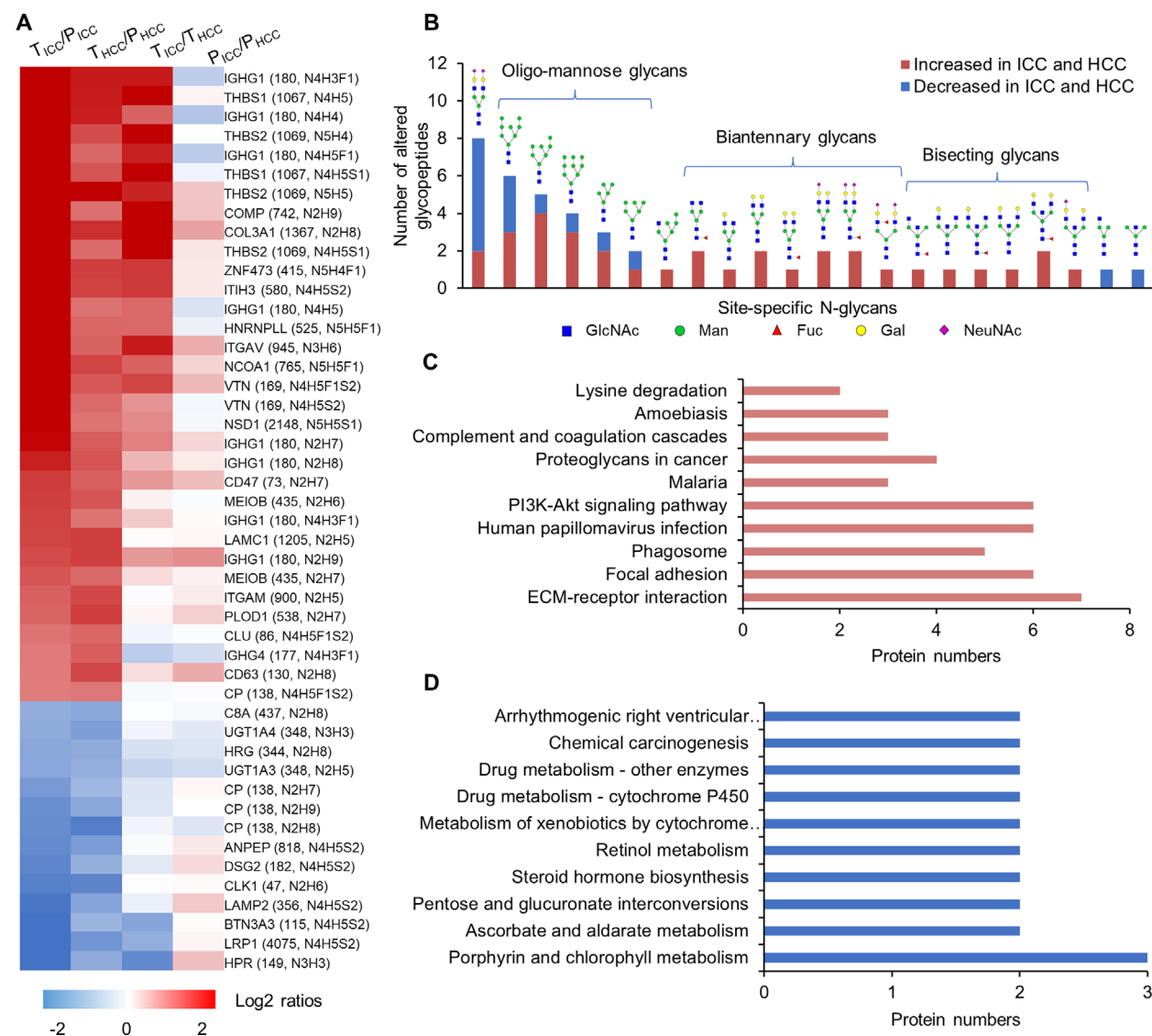
glycopeptides changed at glycosylation level. These results indicated that the site-specific glycosylation changes in ICC were much greater than that of HCC. Altogether, these data indicated a huge distinction between ICC and HCC tumors at both glycoprotein and site-specific glycosylation levels.

### 3.3. Bi-antennary and bisecting glycans are commonly increased in ICC and HCC tumors

To investigate the differences of site-specific glycan occurred at glycosylation level, we first focused on glycopeptides that were commonly changed in both ICC and HCC tumors compared with paired paracancerous tissues. These changes largely represented the common changes of site-specific glycosylation in the tumorigenesis of both ICC and HCC tumors. By excluding the glycopeptides that changed only at protein expression level, we totally identified 47 glycopeptides with at least twofold changes compared with their paracancers, including 33 increased (up to 20-fold increase) and 14 decreased glycopeptides (up to 10-fold decrease) (Fig. 4A and Table S5). These glycopeptides were comprised of 22 N-glycans and 31 glycosites from 32 glycoproteins.

By classifying these changed site-specific glycans based on their attached glycan structures, we found that 13 intact glycopeptides modified by bi-antennary glycans with 0–2 sialic acid, seven glycopeptides modified by bisecting glycans, and 13 glycopeptides modified with oligo-mannose glycans (Man5 to Man9) were increased in both ICC and HCC (Fig. 4B). Their quantification further demonstrated that the greatest alteration occurred in glycopeptides containing bi-antennary and bisecting glycans with 5- to 20-fold changes. These increased glycopeptides were identified from 21 glycoproteins, which were mainly involved in pathways of ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, proteoglycans in cancer, and complement and coagulation cascades (Fig. 4C). As a key regulator involved in cell growth, motility and survival, the aberrant activation of PI3K-Akt signaling pathway could promote the survival and proliferation of tumor cells in many cancers [43]. As an important component of the inflammatory responses, complement and coagulation cascades is involved in tumors including HCC [26]. In addition, the bi-antennary glycans with two sialic acids but no fucose (N4H5S2) and oligo-mannose glycans were also observed in decreased glycans in ICC and HCC (Fig. 4B). These decreased glycopeptides were identified from 12 glycoproteins and three of them (UGT1A3, UGT1A4, CP) were enriched in KEGG pathways including porphyrin and chlorophyll metabolism, ascorbate and aldarate metabolism, retinol metabolism, and drug





**Fig. 4.** Glycopeptides commonly altered in both ICC and HCC tumors compared with paracancer tissues. (A) Heatmap of 33 increased and 14 decreased glycopeptides in both ICC ( $n = 6$ ) and HCC ( $n = 6$ ) tumors. The log<sub>2</sub> ratios between groups were calculated based on the normalized abundance of TMT labeling. T<sub>ICC</sub>, tumors of ICC; P<sub>ICC</sub>, paracancers of ICC; T<sub>HCC</sub>, tumors of HCC; P<sub>HCC</sub>, paracancers of HCC. (B) Frequencies of commonly changed site-specific glycans in ICC ( $n = 6$ ) and HCC ( $n = 6$ ) tumors. (C, D) The KEGG pathways enriched in the corresponding glycoproteins of commonly increased (C) or decreased (D) glycopeptides in ICC and HCC tumors.

metabolism (Fig. 4D). GUT1A3 and UGT1A4 were UDP-glucuronosyltransferases that transforms small lipophilic molecules onto water-soluble metabolites.

Although those glycopeptides were commonly changed in both tumor groups compared with their paired paracancer tissues, varying degrees of change were observed between ICC and HCC. Among the 47 commonly increased glycopeptides, 21 glycopeptides showed additional differences ( $\geq 2$ -fold change) between ICC and HCC but no significant differences were observed between the two groups of their paracancer tissues

(Fig. S3). Those changes were resulted from a greater degree of increase or decrease of glycosylation in ICC than HCC when compared with their paired paracancers.

### 3.4. LacdiNAc-containing glycans, tri-antennary glycans, and core-fucosylated glycans are specifically elevated in ICC tumors

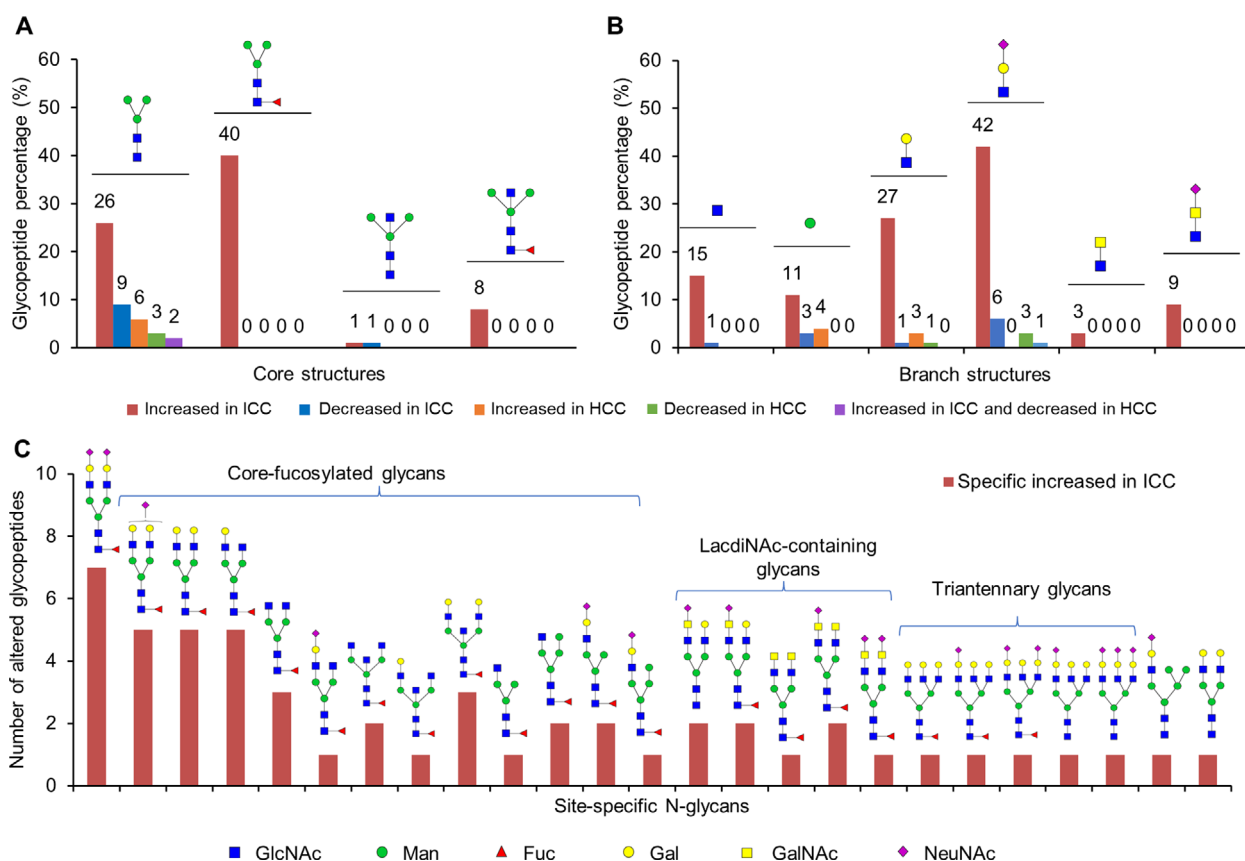
We next studied on the intact glycopeptides that show differential expressions between ICC and HCC tumors ( $\geq 2$ -fold change) but without significant differences

between their paracancer tissues. By excluding the glycopeptides that changed only at protein expression level, we totally identified 95 glycopeptides that were specifically altered (up to 19-fold increase and sixfold decrease) in ICC or HCC tumors compared with their paired paracancers (Fig. S4 and Table S6). These glycopeptides showed additional twofold changes between ICC and HCC tumors and therefore should be specific in two groups of tumors and associated with the differences between ICC and HCC tumors.

Compared with their paired paracancer tissues, the majority of these glycopeptides (84/95, 88.4%) were specifically altered in ICC (with 2- to 19-fold changes), followed by nine glycopeptides specifically altered in HCC (with twofold to fourfold changes) and two glycopeptides with opposite changes in ICC and HCC (Figs S4 and S5). By classifying these glycans based on their core structures, we found that the fucosylated core structure and core structure with both core fucose

and bisecting HexNAc were specifically increased in ICC (Fig. 5A). Moreover, among the six identified branch structures, LacNAc (with or without Neu5Ac) took the majority (69%) of glycopeptides increased in ICC, while LacdiNAc (with or without Neu5Ac) were solely observed in site-specific glycans increased in ICC (Fig. 5B).

Based on the whole glycan structures, we found that eight glycopeptides modified by LacdiNAc-containing (GalNAc $\beta$ 1-4GlcNAc) glycans and four glycopeptides modified with tri-antennary glycans with or without Neu5Ac/fucose were specifically increased in ICC (Fig. 5C). In addition, 47 of those specifically increased glycans in ICC were core-fucosylated. All together, these results indicated that increased glycosylation by core-fucosylated glycans, LacdiNAc-containing glycans, and tri-antennary glycans were uniquely occurred in ICC. The increased levels of core fucosylation and branching are closely related to



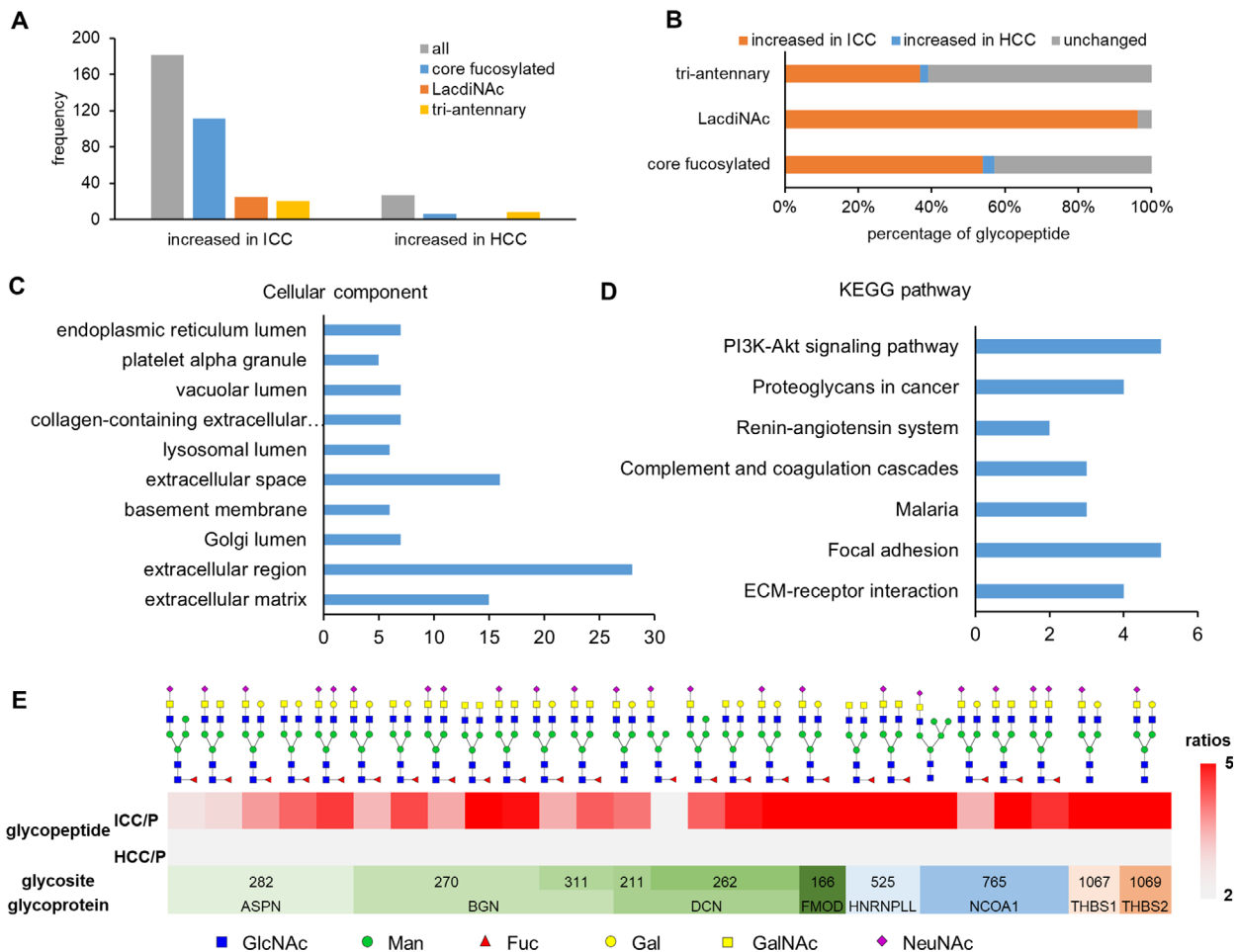
**Fig. 5.** Glycopeptides with altered site-specific glycosylation in either ICC or HCC tumor. (A) Proportions of four types of core structures that uniquely altered in ICC ( $n = 6$ ) and/or HCC ( $n = 6$ ) tumors: solely increased in ICC (in red), solely decreased in ICC (in blue), solely increased in HCC (in orange), solely decreased in HCC (in green), increased in ICC and decreased in HCC (in purple). (B) Identified branch structures and their proportions in the five groups of uniquely altered N-glycopeptides. (C) Frequencies of site-specific glycans that uniquely altered in ICC tumors.

progression of many cancers including HCC [44]; however, increased core fucosylation does not coincide with increased branches in tissues of HCC patient [13]. According to our previously reported results, the core-fucosylated glycans were uniquely increased in HCC with high AFP level while the multi-branched glycans were uniquely increased in HCC with low AFP level [26]. Those increased glycans were identified from 28 glycoproteins, which were involved in pathways of ECM–receptor interaction, complement and coagulation cascades, focal adhesion, proteoglycans in cancer, and PI3K-Akt signaling pathway (Fig. S6). It can be seen that similar pathways were enriched as those commonly altered glycoproteins described above since several key glycoproteins (ITGAV, THBS1, THBS2, VTN, CLU) were modified by both

commonly altered glycans and specifically increased glycans.

### 3.5. Almost all LacdiNAc-containing N-glycans are specifically up-regulated in ICC tumors

To further evaluate the alteration specificities of those glycans in ICC, we investigated all the altered glycopeptides in ICC or HCC at the glycopeptide level. It indicated that altered glycopeptides modified with core-fucosylated glycans or tri-antennary glycans were detected in both ICC and HCC, while all altered glycopeptides with LacdiNAc structures were solely detected in ICC (Fig. 6A). Moreover, among all quantified glycopeptides modified with those glycans, only 57% of core-fucosylated glycans and 39% of tri-



**Fig. 6.** Specific increase of LacdiNAc structures in ICC tumors ( $n = 6$ ). (A) Frequencies of core-fucosylated glycans, LacdiNAc-containing glycans, and tri-antennary glycans uniquely increased in ICC or HCC tumors. (B) Percentages of altered site-specific glycans in ICC or HCC for core-fucosylated, LacdiNAc-containing, and tri-antennary glycans. (C, D) Enriched cellular components (C), and KEGG pathways (D) of the 63 LacdiNAc-containing glycoproteins in ICC and HCC tumors. (E) Heat map of all quantitative LacdiNAc-containing N-glycopeptides showing the fold changes between tumor (ICC or HCC) with paracancer tissues.

**Table 1.** All quantified site-specific glycans with LacdiNAc structures in ICC and HCC. ■: GlcNAc; ●: Man; ▲: Fuc; ●: Gal; ■: GalNAc; ◆: NeuNAc.

No.	Glycoprotein	Gene name	Peptide sequence	Glycan structure	Ratio of glycopeptides			
					ICC/ICCP	HCC/HCCP	ICC/HCC	ICCP/HCCP
1	Asporin	ASPN	ITDIE <sup>282</sup> N <sup>#</sup> GSLANIPR		2.21	1.30	1.56	0.92
2				2.33	1.39	1.69	1.01	
3				3.08	1.46	2.00	0.94	
4				3.75	1.50	2.30	0.92	
5				4.28	1.80	3.04	1.28	
6	Biglycan	BGN	MIE <sup>270</sup> N <sup>#</sup> GSLSFLPTLR		2.73	1.42	2.09	1.08
7				2.85	1.17	2.25	0.93	
8				4.11	1.44	2.51	0.88	
9				3.86	1.29	2.63	0.88	
10				2.91	1.09	2.76	1.04	
11	Decorin	DCN	LLQVYLHSN <sup>311</sup> N <sup>#</sup> ITK		6.02	1.77	2.89	0.85
12				4.83	1.46	3.11	0.95	
13				LGLSFNSISAVD <sup>262</sup> N <sup>#</sup> GSLANTPHLR	1.55	0.78	1.62	0.82
14					3.55	1.17	2.62	0.86
15					3.81	1.14	2.89	0.86
16	Fibromodulin	FMOD	IADT <sup>211</sup> N <sup>#</sup> ITSIPQGLPSPSTELHLDGNK		4.69	1.33	2.95	0.83
17				6.62	1.47	4.00	0.89	
18				7.39	1.75	4.39	1.04	
19	Heterogeneous nuclear ribonucleoprotein L-like	HNRNPLL	LYLDHN <sup>166</sup> N <sup>#</sup> LTR		14.87	1.67	9.85	1.11
20				10.94	1.62	7.42	1.10	
21	Nuclear receptor coactivator 1	NCOA1	STP <sup>765</sup> N <sup>#</sup> LSLDDVK		6.33	1.80	3.53	1.00
22				2.80	1.22	2.54	1.11	
23				7.84	1.32	6.01	1.01	
24				4.39	1.37	4.14	1.29	
25	Thrombospondin-1	THBS1	VV <sup>1067</sup> N <sup>#</sup> STTGPEHLR		9.29	1.86	6.69	1.34
26	Thrombospondin-2	THBS2	VV <sup>1069</sup> N <sup>#</sup> STTGTGEHLR		11.44	1.95	6.63	1.13

antennary glycans were found to be increased in ICC or HCC, while almost all of LacdiNAc-containing glycans (96%) were found to be increased in ICC (Fig. 6B). Overall, the results suggested the increase of LacdiNAc-containing N-glycans is very specific in ICC tumors.

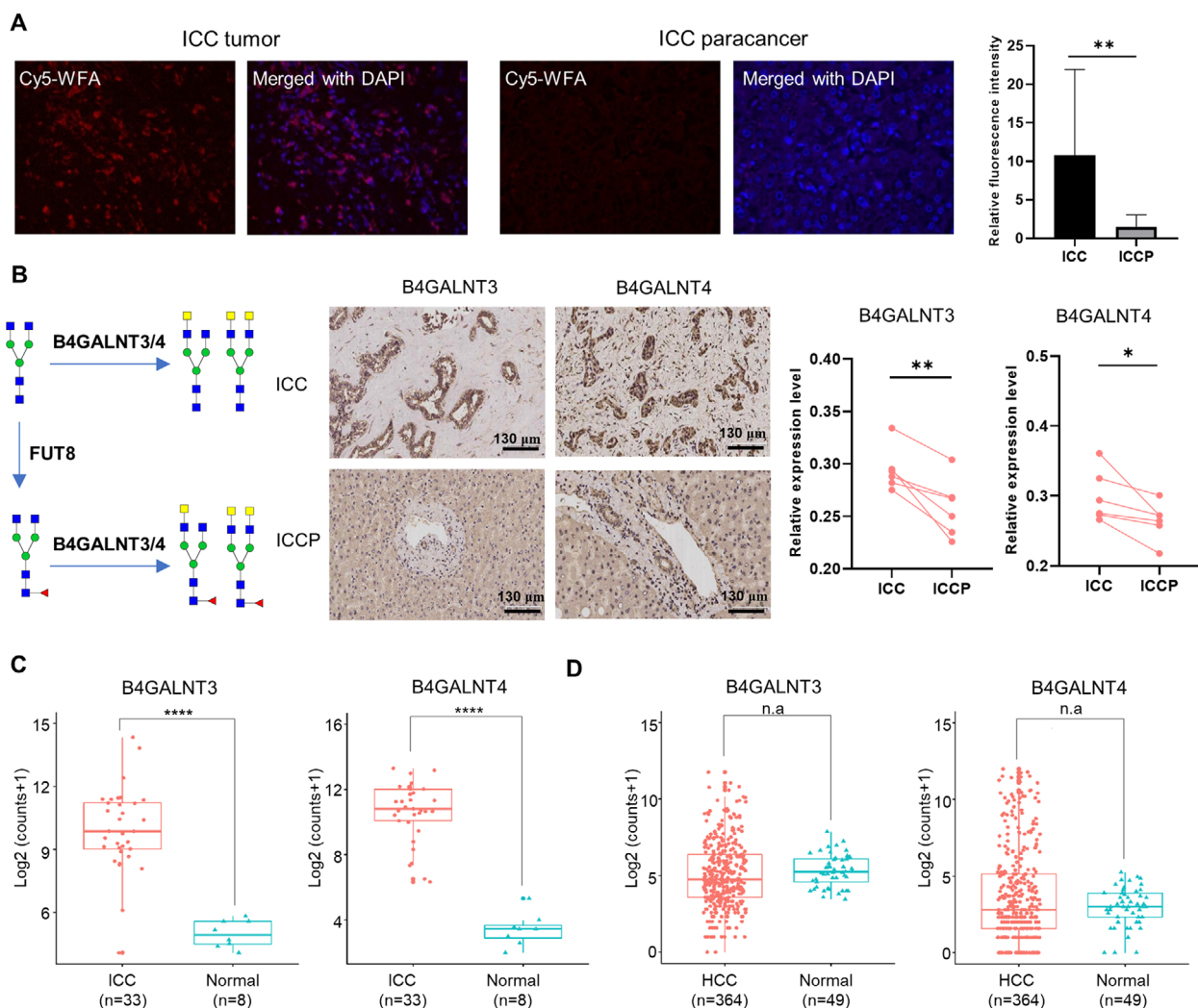
LacdiNAc structures are rarely observed in normal mammalian cells; however, their expressions (evaluated by lectins or glycotransferases) were found to be significantly elevated in tumor tissues of serous ovarian [45,46], peritoneal cancers [47], as well as ICC [48]. Recently, the LacdiNAc-glycosylated prostate-specific antigen was identified as potential serum biomarker for prostate cancer [49]. Despite their importance, their functions are largely unknown mainly limited by available analytical techniques for direct analysis of

LacdiNAc structures. By using our newly developed structural characterization software STRUCGP, we were able to discriminate the LacdiNAc-containing N-glycans from other glycan isomers at the glycopeptide level. A total of 36 N-glycans with solely LacdiNAc (GalNAc $\beta$ 1-4GlcNAc), sialylated LacdiNAc (Neu5Ac $\alpha$ 2-6GalNAc $\beta$ 1-4GlcNAc), and fucosylated LacdiNAc (GalNAc $\beta$ 1-4GlcNAc $\alpha$ 1-3/4Fuc) structures were identified in ICC and HCC. Those LacdiNAc-containing N-glycans were modified on 113 glycopeptides from 33 glycoproteins and were mainly located in extracellular region especially in extracellular matrix, followed by Golgi apparatus part, lysosomal lumen, and endoplasmic reticulum part (Fig. 6C). KEGG analysis showed that they were mainly enriched in ECM-receptor interaction, focal adhesion, complement

and coagulation cascades, proteoglycans in cancer, and PI3K-Akt signaling pathway (Fig. 6D). The cellular component and KEGG analysis showed similar results with that enriched from other glycoproteins as described above, indicating a relative conservative distribution of altered N-glycoproteins in ICC and HCC.

By excluding glycopeptides with less than five PSMs, 26 LacdiNAc-containing glycopeptides were reliably quantified in ICC and HCC. Those glycopeptides comprised of 10 LacdiNAc-containing N-glycans modified at 10 glycosites from eight glycoproteins (ASPN, BGN, DCN, FMOD, HNRNPLL, NCOA1, THBS1,

and THBS2) (Fig. 6E and Table 1). Among them, small leucine-rich proteoglycans (ASPN, BGN, DCN, FMOD) and thrombospondin family proteins (THBS1 and THBS2) have been linked to several tumors including gastric, colon, and bladder cancers [50,51]. NCOA1 is known to promote metastasis and angiogenesis in breast cancer [52], while HNPPL is able to suppress metastasis in colorectal cancer [53]. Furthermore, THBS2 has been proposed as novel diagnostic biomarker for early distal cholangiocarcinoma [54]. However, the glycosylation on them has not been fully explored. In this study, TMT-based quantification of



**Fig. 7.** Confirmation of the increase of LacdiNAc-containing glycans in ICC tumors. (A) Lectin histochemistry analysis of ICC tumors and paired paracancer tissues ( $n = 6$ ,  $\pm$ SD,  $P = 0.0012$  by paired  $t$ -test) based on the WFA binding ( $40\times$  magnification). (B) Immunohistochemistry analysis on the protein expressions of B4GALNT3 ( $n = 6$ ,  $\pm$ SD,  $P = 0.0061$  by paired  $t$ -test) and B4GALNT4 ( $n = 6$ ,  $\pm$ SD,  $P = 0.0341$  by paired  $t$ -test) from ICC tumors and paired paracancer tissues. Scale:  $130\ \mu\text{m}$ . (C) The mRNA levels of B4GALNT3 and B4GALNT4 in ICC tumors ( $n = 33$ ) and normal tissues ( $n = 8$ ) in TCGA database ( $P = 4.5E-9$  for B4GALNT3 and  $P = 1.9E-5$  for B4GALNT4 by Wilcoxon's test). (D) The mRNA levels of B4GALNT3 and B4GALNT4 in HCC tumors ( $n = 364$ ) and normal tissues ( $n = 49$ ) in TCGA database ( $P = 0.13$  for B4GALNT3 and  $P = 0.89$  for B4GALNT4 by Wilcoxon's test).

those glycopeptides revealed that 25 out of the LacdiNAc-containing glycopeptides were elevated in ICC tumor (twofold to fivefold changes) compared with their paired paracancer tissues at the glycopeptide level, while no significant changes were observed in HCC tumor (Fig. 6E). These results not only for the first time confirmed the modification of LacdiNAc-containing N-glycans on these glycoproteins, but also revealed a highly specific alteration of LacdiNAc structures in ICC. This suggested important and specific roles of LacdiNAc-containing glycans and corresponding proteins in ICC, even though their function implications still need further investigations. More importantly, these glycans/glycoproteins/glycopeptides with specific glycosylation alteration have great potential to serve as diagnostic biomarkers or specific targets for therapeutic intervention.

### 3.6. Confirmation of increased LacdiNAc glycans in ICC tumors

To further confirm the increase of LacdiNAc-containing glycans in ICC, we performed lectin histochemistry analysis on six ICC tumors and their paired paracancer tissues using Cy5-labeled wisteria floribunda agglutinin (WFA) for the detection of LacdiNAc structures [55,56]. The histochemical results demonstrated that the binding of WFA was significantly stronger ( $P < 0.01$ ) in ICC tumors compared with their adjacent normal tissues (Fig. 7A). The enhancement of WFA-reactive glycans in ICC tumors was in agreement with previous reported histochemical results of ICC using biotinylated WFA [17]. The general increase of LacdiNAc structures in ICC tumor was also confirmed by the elevated protein expression of corresponding  $\beta$ 1-4-acetylgalactosamine transferases B4GALNT3 ( $n = 6$ ,  $P < 0.01$ ) and B4GALNT4 ( $n = 6$ ,  $P < 0.05$ ) [45,57] measured by the immunohistochemistry assay (Fig. 7B). Consistently, based on the transcriptome data in the TCGA database, the mRNA levels of B4GALNT3 and B4GALNT4 were also significantly elevated ( $P < 0.0001$ ) in ICC tumors ( $n = 33$ ) compared with normal tissues ( $n = 8$ ) (Fig. 7C), while no significant difference was observed in HCC tumors ( $n = 364$ ) compared with the corresponding normal tissues ( $n = 49$ ) (Fig. 7D). Altogether, the above results further confirmed that the LacdiNAc structure was specifically increased in ICC tumors.

## 4. Conclusion

By using our newly developed structural characterization strategy and software STRUCGP, we comprehensively

identified site-specific N-glycan structures at glycopeptide level in ICC and HCC tissues. Specifically, site-specific N-glycans with uncommon LacdiNAc (GalNAc $\beta$ 1-4GlcNAc) structures were distinguished from their isomeric glycans. Integrating with TMT-based quantification, we uncovered the great differences of site-specific N-glycans between tumors (ICC or HCC) and non-tumors as well as between ICC and HCC tumors. Notably, the majority of LacdiNAc-containing glycans were specifically increased in ICC tumors, which were further confirmed by lectin histochemistry and the high expression of  $\beta$ 1-4GalNAc transferases in ICC at both protein and mRNA levels. Further validation and functional study, which are still underway in our laboratory, will allow us to uncover their functional implications in ICC.

Site-specific glycans have been recognized as important diagnostic biomarkers and therapeutic targets for many diseases including cancers. However, it is highly challenging to precisely characterize glycosylation of proteins. By direct structural characterization of intact N-glycopeptides, our approach not only identifies functional glycan structures but also identifies the modified glycoproteins. The identified functional glycan structures can be further pursued as targets for specific biomarkers of detection and treatment, and at the meantime, the corresponding glycoproteins allows us to perform further functional investigation. By discovering specific alteration of LacdiNAc structures and modified proteins, we provide a new perspective for distinguishing ICC from HCC and exploring its mechanism. All in all, structural characterization of site-specific N-glycans has great potential to identify novel aberrant glycosylation in cancers, which provides the basis on their biomarker discovery and functional investigation.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (no. 81773180, 91853123, 32101192, 21705127, and 81800655), National Key Research and Development Program of China (no. 2019YFA09005200), and China Postdoctoral Science Foundation (no. 2019TQ0260 and 2019M663798).

## Conflict of interest

The authors declare no conflict of interest.

## Author contributions

JL, TZ, and SS designed the experiments; FM and LH collected the tissue samples and clinical information;

TZ performed experiments with help from LJ, LD, and BZ; JL analyzed data with the help from JingL, JS, CM, and DL; JL and SS wrote and edited the manuscript; and all the authors contributed to manuscript review.

## Peer review

The peer review history for this article is available at <https://publons.com/publon/10.1002/1878-0261.13147>.

## Data accessibility

The mass spectrometry data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [58] with the dataset identifier [PXD020277](https://doi.org/10.1002/1878-0261.13147).

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;**68**:394–424. <https://doi.org/10.3322/caac.21492>
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;**71**:209–49. <https://doi.org/10.3322/caac.21660>
- Shaib YH, El-Serag HB, Davila JA, Morgan R, McGlynn KA. Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. *Gastroenterology*. 2005;**128**:620–6. <https://doi.org/10.1053/j.gastro.2004.12.048>
- Seehawer M, Heinzmann F, D'Artista L, Harbig J, Roux PF, Hoenicke L, et al. Necroptosis microenvironment directs lineage commitment in liver cancer. *Nature*. 2018;**562**:69–75. <https://doi.org/10.1038/s41586-018-0519-y>
- Kim SJ, Lee JM, Han JK, Kim KH, Lee JY, Choi BI. Peripheral mass – forming cholangiocarcinoma in cirrhotic liver. *Am J Roentgenol*. 2007;**189**:1428–34. <https://doi.org/10.2214/ajr.07.2484>
- Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Cancer Control*. 2017;**24**:1073274817729245. <https://doi.org/10.1177/1073274817729245>
- Sia D, Tovar V, Moeini A, Llovet JM. Intrahepatic cholangiocarcinoma: pathogenesis and rationale for molecular therapies. *Oncogene*. 2013;**32**:4861–70. <https://doi.org/10.1038/onc.2012.617>
- Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochem Biophys Acta*. 1999;**1473**:4–8. [https://doi.org/10.1016/s0304-4165\(99\)00165-8](https://doi.org/10.1016/s0304-4165(99)00165-8)
- Trefts E, Gannon M, Wasserman DH. The liver. *Curr Biol*. 2017;**27**:R1147–51. <https://doi.org/10.1016/j.cub.2017.09.019>
- Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, et al. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res*. 2009;**8**:651–61. <https://doi.org/10.1021/pr8008012>
- de Oliveira RM, Ornelas Ricart CA, Araujo Martins AM. Use of mass spectrometry to screen glycan early markers in hepatocellular carcinoma. *Front Oncol*. 2018;**7**:328. <https://doi.org/10.3389/fonc.2017.00328>
- Dall'Olio F, Chiricolo M, D'Errico A, Gruppioni E, Altimari A, Fiorentino M, et al. Expression of beta-galactoside alpha 2,6 sialyltransferase and of alpha 2,6-sialylated glycoconjugates in normal human liver, hepatocarcinoma, and cirrhosis. *Glycobiology*. 2004;**14**:39–49. <https://doi.org/10.1093/glycob/cwh002>
- Mehta A, Norton P, Liang H, Comunale MA, Wang M, Rodemich-Betesh L, et al. Increased levels of tetra-antennary N-linked glycan but not core fucosylation are associated with hepatocellular carcinoma tissue. *Cancer Epidemiol Biomarkers Prev*. 2012;**21**:925–33. <https://doi.org/10.1158/1055-9965.EPI-11-1183>
- Li D, Mallory T, Satomura S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin Chim Acta*. 2001;**313**:15–9. [https://doi.org/10.1016/s0009-8981\(01\)00644-1](https://doi.org/10.1016/s0009-8981(01)00644-1)
- Indramanee S, Silsirivanit A, Pairojkul C, Wongkham C, Wongkham S. Aberrant glycosylation in cholangiocarcinoma demonstrated by lectin-histochemistry. *Asian Pac J Cancer Prev*. 2012;**13**:119–24. <https://doi.org/10.7314/APJCP.2012.13.KKSuppl.119>
- Matsuda A, Kuno A, Matsuzaki H, Kawamoto T, Shikanai T, Nakanuma Y, et al. Glycoproteomics-based cancer marker discovery adopting dual enrichment with Wisteria floribunda agglutinin for high specific glyco-diagnosis of cholangiocarcinoma. *J Proteomics*. 2013;**85**:1–11. <https://doi.org/10.1016/j.jprot.2013.04.017>
- Shoda J, Matsuda A, Shida T, Yamamoto M, Nagino M, Tsuyuguchi T, et al. Wisteria floribunda agglutinin-sialylated mucin core polypeptide 1 is a sensitive biomarker for biliary tract carcinoma and intrahepatic cholangiocarcinoma: a multicenter study. *J Gastroenterol*. 2017;**52**:218–28. <https://doi.org/10.1007/s00535-016-1230-0>

- 18 Powers TW, Holst S, Wuhrer M, Mehta AS, Drake RR. Two-dimensional N-glycan distribution mapping of hepatocellular carcinoma tissues by MALDI-imaging mass spectrometry. *Biomolecules*. 2015;**5**:2554–72. <https://doi.org/10.3390/biom5042554>
- 19 West CA, Wang M, Herrera H, Liang H, Black A, Angel PM, et al. N-linked glycan branching and fucosylation are increased directly in Hcc tissue as determined through in situ glycan imaging. *J Proteome Res*. 2018;**17**:3454–62. <https://doi.org/10.1021/acs.jproteome.8b00323>
- 20 Cheng L, Gao S, Song X, Dong W, Zhou H, Zhao L, et al. Comprehensive N-glycan profiles of hepatocellular carcinoma reveal association of fucosylation with tumor progression and regulation of FUT8 by microRNAs. *Oncotarget*. 2016;**7**:61199–214. <https://doi.org/10.18632/oncotarget.11284>
- 21 Han Y, Xiao K, Tian Z. Comparative glycomics study of cell-surface N-glycomes of HepG2 versus LO2 cell lines. *J Proteome Res*. 2019;**18**:372–9. <https://doi.org/10.1021/acs.jproteome.8b00655>
- 22 Xiao K, Han Y, Tian Z. Large-scale identification and visualization of human liver N-glycome enriched from LO2 cells. *Anal Bioanal Chem*. 2018;**410**:4195–202. <https://doi.org/10.1007/s00216-018-1070-2>
- 23 Nie H, Liu X, Zhang Y, Li T, Zhan C, Huo W, et al. Specific N-glycans of hepatocellular carcinoma cell surface and the abnormal increase of core-alpha-1, 6-fucosylated triantennary glycan via N-acetylglucosaminyltransferases-IVa regulation. *Sci Rep*. 2015;**5**:16007. <https://doi.org/10.1038/srep16007>
- 24 Qin H, Chen Y, Mao J, Cheng K, Sun D, Dong M, et al. Proteomics analysis of site-specific glycoforms by a virtual multistage mass spectrometry method. *Anal Chim Acta*. 2019;**1070**:60–8. <https://doi.org/10.1016/j.aca.2019.04.025>
- 25 Xiao K, Tian Z. GPSeeker enables quantitative structural N-glycoproteomics for site- and structure-specific characterization of differentially expressed N-glycosylation in hepatocellular carcinoma. *J Proteome Res*. 2019;**18**:2885–95. <https://doi.org/10.1021/acs.jproteome.9b00191>
- 26 Zhao T, Jia L, Li J, Ma C, Wu J, Shen J, et al. Heterogeneities of site-specific N-glycosylation in HCC tumors with low and high AFP concentrations. *Front Oncol*. 2020;**10**:496. <https://doi.org/10.3389/fonc.2020.00496>
- 27 Shen J, Jia L, Dang L, Su Y, Zhang J, Xu Y, et al. StrucGP: de novo structural sequencing of site-specific N-Glycan on glycoproteins using a modularization strategy. *Nat Methods*. 2021;**18**:921–9. <https://doi.org/10.1038/s41592-021-01209-0>
- 28 Jia L, Li J, Li P, Liu D, Li J, Shen J, et al. Site-specific glycoproteomic analysis revealing increased core-fucosylation on FOLR1 enhances folate uptake capacity of HCC cells to promote EMT. *Theranostics*. 2021;**11**:6905–21. <https://doi.org/10.7150/thno.56882>
- 29 Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003;**75**:1895–904. <https://doi.org/10.1021/ac0262560>
- 30 Yang W, Shah P, Hu Y, Toghi Eshghi S, Sun S, Liu Y, et al. Comparison of enrichment methods for intact N- and O-linked glycopeptides using strong anion exchange and hydrophilic interaction liquid chromatography. *Anal Chem*. 2017;**89**:11193–7. <https://doi.org/10.1021/acs.analchem.7b03641>
- 31 Keller A, Eng J, Zhang N, Li X-J, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005;**1**:2005.0017.
- 32 Liu MQ, Zeng WF, Fang P, Cao WQ, Liu C, Yan GQ, et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun*. 2017;**8**:438. <https://doi.org/10.1038/s41467-017-00535-2>
- 33 Toghi Eshghi S, Shah P, Yang W, Li X, Zhang H. GPQuest: a spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides. *Anal Chem*. 2015;**87**:5181–8. <https://doi.org/10.1021/acs.analchem.5b00024>
- 34 Le S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;**25**:1–18.
- 35 Qin Y, Zhong Y, Ma T, Wu F, Wu H, Yu H, et al. Alteration of liver glycopatterns during cirrhosis and tumor progression induced by HBV. *Glycoconj J*. 2016;**33**:125–36. <https://doi.org/10.1007/s10719-015-9645-z>
- 36 Farshidfar F, Zheng S, Gingras M-C, Newton Y, Shih J, Robertson AG, et al. Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep*. 2017;**18**:2780–94. <https://doi.org/10.1016/j.celrep.2017.02.033>
- 37 Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, et al. Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *Ebiomedicine*. 2019;**40**:457–70. <https://doi.org/10.1016/j.ebiom.2018.12.058>
- 38 Wang W, Kaluza A, Nouta J, Nicolardi S, Ferens-Sieczkowska M, Wuhrer M, et al. High-throughput glycopeptide profiling of prostate-specific antigen from seminal plasma by MALDI-MS. *Talanta*. 2021;**222**:121495. <https://doi.org/10.1016/j.talanta.2020.121495>
- 39 Sanda M, Morrison L, Goldman R. N- and O-glycosylation of the SARS-CoV-2 spike protein. *Anal Chem*. 2021;**93**:2003–9. <https://doi.org/10.1021/acs.analchem.0c03173>



- 40 Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell Mol Biol Lett*. 2019;**24**:38. <https://doi.org/10.1186/s11658-019-0162-0>
- 41 Eke I, Cordes N. Focal adhesion signaling and therapy resistance in cancer. *Semin Cancer Biol*. 2015;**31**:65–75. <https://doi.org/10.1016/j.semcancer.2014.07.009>
- 42 Iozzo RV, Sanderson RD. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J Cell Mol Med*. 2011;**15**:1013–31. <https://doi.org/10.1111/j.1582-4934.2010.01236.x>
- 43 Jiang N, Dai Q, Su X, Fu J, Feng X, Peng J. Role of PI3K/AKT pathway in cancer: the framework of malignant behavior. *Mol Biol Rep*. 2020;**47**:4587–629. <https://doi.org/10.1007/s11033-020-05435-1>
- 44 Miyoshi E, Moriwaki K, Nakagawa T. Biological function of fucosylation in cancer biology. *J Biochem*. 2008;**143**:725–9. <https://doi.org/10.1093/jb/mvn011>
- 45 Anugraham M, Jacob F, Everest-Dass AV, Schoetzau A, Nixdorf S, Hacker NF, et al. Tissue glycomics distinguish tumour sites in women with advanced serous adenocarcinoma. *Mol Oncol*. 2017;**11**:1595–615. <https://doi.org/10.1002/1878-0261.12134>
- 46 Machado E, Kandzia S, Carilho R, Altevogt P, Conradt HS, Costa J. N-Glycosylation of total cellular glycoproteins from the human ovarian carcinoma SKOV3 cell line and of recombinantly expressed human erythropoietin. *Glycobiology*. 2011;**21**:376–86. <https://doi.org/10.1093/glycob/cwq170>
- 47 Pirro M, Rombouts Y, Stella A, Neyrolles O, Burlet-Schiltz O, van Vliet SJ, et al. Characterization of macrophage galactose-type lectin (MGL) ligands in colorectal cancer cell lines. *Biochim Biophys Acta Gen Subj*. 2020;**1864**:129513. <https://doi.org/10.1016/j.bbagen.2020.129513>
- 48 Matsuda A, Kuno A, Kawamoto T, Matsuzaki H, Irimura T, Ikehara Y, et al. *Wisteria floribunda* agglutinin-positive mucin 1 is a sensitive biliary marker for human cholangiocarcinoma. *Hepatology*. 2010;**52**:174–82. <https://doi.org/10.1002/hep.23654>
- 49 Haga Y, Uemura M, Baba S, Inamura K, Takeuchi K, Nonomura N, et al. Identification of multisialylated LacdiNAc structures as highly prostate cancer specific glycan signatures on PSA. *Anal Chem*. 2019;**91**:2247–54. <https://doi.org/10.1021/acs.analchem.8b04829>
- 50 Appunni S, Anand V, Khandelwal M, Gupta N, Rubens M, Sharma A. Small leucine rich proteoglycans (decorin, biglycan and lumican) in cancer. *Clin Chim Acta*. 2019;**491**:1–7. <https://doi.org/10.1016/j.cca.2019.01.003>
- 51 Wang P, Zeng Z, Lin C, Wang J, Xu W, Ma W, et al. Thrombospondin-1 as a potential therapeutic target: multiple roles in cancers. *Curr Pharm Des*. 2020;**26**:2116–36. <https://doi.org/10.2174/1381612826666200128091506>
- 52 Qin L, Xu Y, Xu Y, Ma G, Liao L, Wu Y, et al. NCOA1 promotes angiogenesis in breast tumors by simultaneously enhancing both HIF1 alpha- and AP-1-mediated VEGFa transcription. *Oncotarget*. 2015;**6**:23890–904. <https://doi.org/10.18632/oncotarget.4341>
- 53 Sakuma K, Sasaki E, Kimura K, Komori K, Shimizu Y, Yatabe Y, et al. HNRNPLL, a newly identified colorectal cancer metastasis suppressor, modulates alternative splicing of CD44 during epithelial-mesenchymal transition. *Gut*. 2018;**67**:1103–11. <https://doi.org/10.1136/gutjnl-2016-312927>
- 54 Byrling J, Hilmersson KS, Ansari D, Andersson R, Andersson B. Thrombospondin-2 as a diagnostic biomarker for distal cholangiocarcinoma and pancreatic ductal adenocarcinoma. *Clin Transl Oncol*. 2021. <https://doi.org/10.1007/s12094-021-02685-8>
- 55 Fukushima K, Satoh T, Baba S, Yamashita K. alpha1,2-Fucosylated and beta-N-acetylgalactosaminylated prostate-specific antigen as an efficient marker of prostatic cancer. *Glycobiology*. 2010;**20**:452–60. <https://doi.org/10.1093/glycob/cwp197>
- 56 Kaya T, Kaneko T, Kojima S, Nakamura Y, Ide Y, Ishida K, et al. High-sensitivity immunoassay with surface plasmon field-enhanced fluorescence spectroscopy using a plastic sensor chip: application to quantitative analysis of total prostate-specific antigen and GaINAc beta 1–4GlcNAc-linked prostate-specific antigen for prostate cancer diagnosis. *Anal Chem*. 2015;**87**:1797–803. <https://doi.org/10.1021/ac503735e>
- 57 Gotoh M, Sato T, Kiyohara K, Kameyama A, Kikuchi N, Kwon YD, et al. Molecular cloning and characterization of beta 1,4-N-acetylgalactosaminyltransferases IV synthesizing N, N'-diacetyllactosediamine. *FEBS Lett*. 2004;**562**:134–40. [https://doi.org/10.1016/s0014-5793\(04\)00219-4](https://doi.org/10.1016/s0014-5793(04)00219-4)
- 58 Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*. 2013;**41**:D1063–9.

## Supporting information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** LC chromatograph and MS/MS spectra of glycan isoforms identified using StrucGP.

**Fig. S2.** Enriched cellular component and KEGG pathways of the altered glycoproteins in ICC and HCC tumors.

**Fig. S3.** Commonly altered glycopeptides in both ICC and HCC tumors compared with paracancer tissues

with additional differences between ICC and HCC tumors.

**Fig. S4.** Heatmap of the 95 glycopeptides specially altered in ICC or HCC tumors.

**Fig. S5.** Specifically altered glycopeptides in ICC or HCC tumors.

**Fig. S6.** KEGG pathway analyses of specifically altered glycopeptides in ICC tumors.

**Table S1.** The clinical information of ICC and HCC patients involved in this study.

**Table S2.** All identified intact glycopeptides in tumors and paracancerous samples of ICC and HCC.

**Table S3.** All quantified intact glycopeptides in tumors and paracancerous samples of ICC and HCC.

**Table S4.** Altered glycoproteins in ICC or HCC tumors compared with paired paracancerous tissues.

**Table S5.** Commonly altered glycopeptides in ICC and HCC tumors.

**Table S6.** Differentially altered glycopeptides in ICC and HCC tumors.