Journal of proteome research

Article

pubs.acs.org/jpr

Comprehensive Mapping of Protein N-Glycosylation in Human Liver by Combining Hydrophilic Interaction Chromatography and Hydrazide Chemistry

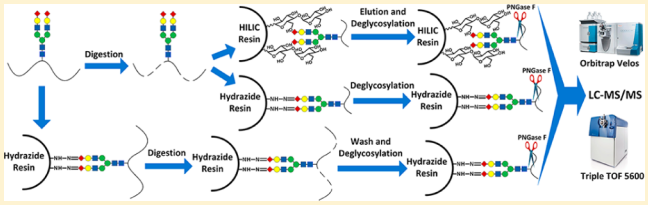
Jun Zhu,[†,§] Zhen Sun,[†,§] Kai Cheng,[†] Rui Chen,[†] Mingliang Ye,[†] Bo Xu,[†] Deguang Sun,[‡] Liming Wang,[‡] Jing Liu,[†] Fangjun Wang,*[,†] and Hanfa Zou*[,†]

[†]Key Lab of Separation Science for Analytical Chemistry, National Chromatographic Research and Analysis Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, China

[‡]The Second Affiliated Hospital of Dalian Medical University, 467 Zhongshan Road, Dalian 116027, China

Supporting Information

ABSTRACT: Although glycoproteomics is greatly developed in recent years, our knowledge about N-glycoproteome of human tissues is still very limited. In this study, we comprehensively mapped the N-glycosylation sites of human liver by combining click maltose−hydrophilic interaction chromatography (HILIC) and the improved hydrazide chemistry. The specificity could be as high as 90% for hydrazide chemistry and 80% for HILIC. Altogether, we identified 14 480 N-glycopeptides matched with N-!P-[S|T|C] sequence motif from human liver, corresponding to 2210 N-glycoproteins and 4783 N-glycosylation sites. These N-glycoproteins are widely involved into different types of biological processes, such as hepatic stellate cell activation and acute phase response of human liver, which all highly associate with the progression of liver diseases. Moreover, the exact N-glycosylation sites of some key-regulating proteins within different human liver physiological processes were also obtained, such as E-cadherin, transforming growth factor beta receptor and 29 members of G protein coupled receptors family.



KEYWORDS: *N-glycosylation sites, N-glycoproteins, HILIC, hydrazide chemistry, human liver*

INTRODUCTION

Glycosylation has long been recognized as one of the most common post-translational modifications (PTMs) of protein, and over 50% of all proteins in mammalian are estimated to undergo glycosylation during their lifetime.[1] Among all types of glycosylation, N-glycosylation is widely distributed in organisms with the β-glycosylamine linkage of GlcNAc to asparagine, which is involved in many molecular and cellular processes, such as ion transport, cell communication, and signaling transduction.[2−4] Aberrant N-glycosylation has been proved to be related to carcinogenesis and tumor progression.[5] Many clinical used cancer biomarkers are N-glycoproteins, such as prostate-specific antigen (PSA) for prostate cancer, alpha-fetoprotein (AFP) for hepatocellular carcinoma, and HER2/neu for breast cancer.[6] Given the biological importance of this modification, the N-glycoproteome of varieties of samples, such as liver, brain, heart and serum/plasma, were extensively investigated in the past decade.[6−11] However, the majority of currently known N-glycosylation sites are originated from mouse/rat tissues and human cell lines.

High-efficient enrichment of N-glycoproteins or N-glycopeptides before MS detection is usually the first step of N-glycoproteome analysis due to the much lower stoichiometry of N-glycoproteins compared with other unmodified proteins in complex biological samples. Different enrichment methods have been developed in recent years, that is, lectin affinity chromatography, hydrazide chemistry and hydrophilic interaction chromatography (HILIC), and so on. Zielinska et al. developed a filter-aided sample preparation (FASP)-based strategy in which multiple lectins were applied for the enrichment of N-glycopeptides from four mouse tissues and plasma.[9] Altogether, they identified more than 5000 N-glycosylation sites in 2352 N-glycoproteins. Later, this technique was also applied to seven nonmammalian systems, and hundreds to thousands of N-glycosylation sites were identified from different species.[10] Apart from lectin affinity chromatography, hydrazide chemistry and HILIC were both widely utilized in N-glycoproteomics. Chen et al. combined multienzyme digestion and hydrazide chemistry for human liver N-glycoproteome analysis with an identification of 939 N-glycosylation sites and 523 N-glycoproteins.[12] Parker et al. applied hydrazide chemistry, titanium dioxide, and ZIC−HILIC to increase the glycoproteome coverage of rat heart and identified 1556 N-glycosylation sites corresponding to 972 N-glycoproteins.[13] Recently, we demonstrated that the click maltose-HILIC has good efficiency and selectivity for the enrichment of N-glycopeptides.[14]

Received: December 8, 2013
Published: February 5, 2014

dx.doi.org/10.1021/pr401200h | J. Proteome Res. 2014, 13, 1713−1721

Liver is the largest organ in human body and plays a critical role in energy metabolism, protein synthesis, internal environment homeostasis, blood proteins secretion, detoxification in the adult, and hematopoiesis in the embryo. Liver cancer is the third most fatal cancer after the lung and the stomach carcinomas.[15,16] Because of the great needs of the comprehensive protein atlas of human liver, the Human Liver Proteome Project (HLPP) was launched in 2002 to uncover the proteomic basis of liver development, physiology, and pathology that is helpful in developing liver-specific diagnostics and therapeutics.[17] Chinese Human Liver Proteome Profiling Consortium already announced the proteome data set of 6788 identified proteins and the transcriptome data set of 11 205 expressed genes for human liver in 2009.[18,19] However, the largest N-glycoproteome data set of human liver contained only 939 confident N-glycosylation sites.[12] Therefore, comprehensive mapping of the N-glycoproteome of human liver is still urgently needed.

To increase the coverage of human liver N-glycoproteome, we combined both hydrazide chemistry and HILIC for the enrichment of human liver N-glycoproteins or N-glycopeptides and the multienzyme digestion, and two different types of mass spectrometers were further adopted. Finally, 14 480 N-glycopeptides corresponding to 2210 N-glycoproteins and 4783 N-glycosylation sites were confidently identified in total from human liver. It is the largest data set of human N-glycoproteome and also the largest data set of N-glycoproteome from a single mammalian tissue to date.

EXPERIMENTAL SECTION

Chemicals and Materials

Trypsin, Glu-C, and chymotrypsin were purchased from Sigma (St. Louis, MO). PNGase F was from New England Biolabs (Ipswich, MA). Chemical reagents of iodacetamide (IAA), 1,4-dithiothreitol (DTT), and trifluoroacetic acid (TFA) were obtained from Sigma (St. Louis, MO). Formic acid (FA) was obtained from Fluka (Buches, Germany). Acetonitrile (ACN, HPLC grade) was from Merck (Darmstadt, Germany). Ammonium bicarbonate and urea were from Bio Basic (Ontario, Canada). Pure water used in all experiments was purified with a Milli-Q system (Millipore, Milford, MA). Other chemicals used were either of analytical grade or better.

The centrifugal filter units (Amicon Ultra-0.5 mL) were purchased from Millipore (Milford, MA). GELoader tips (20 μL) were purchased from Eppendorf (Hamburg, Germany). C18 AQ beads (3 and 5 μm, 120 Å) were from Michrom BioResources (Auburn, CA). Click maltose-HILIC beads (4 μm, 100 Å) were kindly provided by Prof. Xinmiao Liang (Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian, China). Hydrazide beads were obtained from Bio-Rad (Hercules, CA).

Protein Extraction

The human liver was the noncancerous liver tissues ≥2 cm outside the hepatic cancer nodules removed by surgical operation from patients at the Second Affiliated Hospital of Dalian Medical University (Dalian, China). The noncancerous liver tissue has been verified by histopathological examination, which excluded the presence of invading or microscopic metastatic cancer cells. The utilization of human tissues complied with guidelines of Ethics Committee of the Hospital. The noncancerous liver tissues from four patients with hepatic cancer (two male and two female) were pooled together for all experiments.

The isolated human liver tissues were first cut into pieces and washed several times with PBS buffer to remove the existing blood. Then, the liver tissues were placed in an ice-cold homogenization buffer I that consisted of 10 mM HEPES, 1.5 mM $MgCl_2$, 5 mM KCl, 0.1 mM EDTA, and 2% protease inhibitor cocktail (pH 7.4), followed by homogenization using an IKA Ultra Turbax blender. After that, the samples were transferred to a Potter–Elvejhem homogenizer with a Teflon piston and homogenized for a second time on ice. The homogenates were then centrifuged at 1000g for 5 min to pellet the nuclei and debris. The supernatant was collected, and five volumes of buffer II (0.1 M Tris-HCl, 4% SDS, pH 7.4) were added for sonication using an ultrasonic cell disrupter (3 s with 3 s intervals for 180 times at 400 W) and centrifuged at 20 000g for 15 min. The supernatant was collected, and the concentration of proteins was determined by BCA assay.

Protein Digestion

Three batches of 1 to 2 mg human liver proteins were first added to three centrifugal filter units (Amicon Ultra-0.5) with 10 KDa cutoff, respectively. Then, 300 μL of 8 M urea/100 mM NH_4HCO_3 was added, followed by reduction with DTT and alkylation with IAA, as we previously reported.[20] The samples were desalted by centrifugation for 15 min at 14 000 g and washed with 400 μL of 100 mM NH_4HCO_3 three times. After that, three different enzymes (trypsin, trypsin and Glu-C, and chymotrypsin) were added to the three filter units for digestion, respectively. The enzyme-to-protein ratio of each enzyme was as follows: trypsin-to-protein ratio was 1:20; trypsin and Glu-C were both added at enzyme-to-protein ratio of 1:20; and chymotrypsin-to-protein ratio was 1:10. The three batches of samples were all incubated at 37 °C overnight. Finally, the protein digest was collected by centrifugation and lyophilized to dryness in a Speed Vac (Thermo, USA).

Enrichment of N-Glycopeptides with Centrifugation-Assisted Click Maltose–HILIC

Click maltose–HILIC tips were made as previously described.[14] In brief, the GELoader tip was first packed with a small piece of cotton wool as the sieve. Then, 20 mg of click maltose-HILIC beads (4 μm, 100 Å) was packed into the tip by centrifugation at 4000g for 10 min. Finally, the tip was washed by 20 μL of H_2O and equilibrated with 20 μL of 80% ACN before usage.

The protein sample was first redissolved in the enrichment loading buffer (80% ACN/1% TFA) and then pipetted into a HILIC tip. After centrifugation at 4000g for ∼15 min, the HILIC tip was washed with 40 μL of loading buffer three times. Finally, the enriched N-glycopeptides were eluted with 80 μL of H_2O and lyophilized to dryness. For deglycosylation, 200 units of PNGase F in 50 μL of 40 mM NH_4HCO_3 was added and incubated at 37 °C overnight.

Enrichment of N-Glycopeptides and Proteins with Hydrazide Chemistry

For enrichment of N-glycopeptides, the protein samples were first digested with multienzymes. The dried protein digest was reconstituted in oxidation buffer (150 mM NaCl, 100 mM NaAc, 2% SDS, 2% Triton, pH 5.5), and 10 mM final concentration of $NaIO_4$ was added. The reaction was kept in the dark for 1 h and quenched by the addition of 20 mM $Na_2S_2O_3$. Then, the oxidized glycosylated peptides were

dx.doi.org/10.1021/pr401200h | J. Proteome Res. 2014, 13, 1713–1721

coupled to the prewashed hydrazide resins (Bio-Rad, USA) overnight at room temperature.

For enrichment of N-glycoproteins, the protein samples were first desalted with spin desalting column after incubating in boiling water for 10 min. Then, the desalted solutions were resuspended with oxidation buffer. 50 mM $NaIO_4$ was added to the solutions, and the reactions were kept in dark at room temperature for 1 h. Finally, the oxidized solutions were added to the prewashed hydrazide resins after 100 mM $Na_2S_2O_3$ was added to quench the oxidation reaction. The coupling reaction was carried out with gentle shaking at room temperature overnight. After the removal of supernatant from the solution of coupling reaction, the captured N-glycoproteins were digested on the resins with multienzymes (trypsin, trypsin and Glu-C, chymotrypsin). The digestion conditions were the same as those previously mentioned.

After digestion, the hydrazide beads were washed with 1.5 M NaCl, 80% ACN, and 100 mM NH_4HCO_3 sequentially. Then, 1000 units of PNGase F in 10 mM NH_4HCO_3 were added to the resins to release the N-glycopeptides. The released deglycosylated peptides were carefully collected by gentle centrifugation.

Chemical Deamidation Investigation

To test the chemical deamidation in digestion, enrichment, and deglycosylation processes, we incubated the same liver proteins digested and enriched as previously described at 37 °C overnight without PNGase F in NH_4HCO_3 buffer. The samples were collected and lyophilized to dryness for analysis.

Mass Spectrometry Analysis

The enriched N-glycopeptides were analyzed by both LTQ-Orbitrap Velos (Thermo, USA) and Triple-TOF 5600 (AB SCIEX, USA). The LTQ-Orbitrap Velos system was equipped with an Accela 600 HPLC (Thermo, San Jose, CA) involving a 3 cm C18 capillary trap column (200 μm i.d.) and a 12 cm C18 capillary analysis column (75 μm i.d.) with spray emitter as we previously reported.[20] The trap column was packed with C18 AQ beads (5 μm, 120 Å), and the separation column was packed with C18 AQ beads (3 μm, 120 Å).[20,21] The deglycosylated peptides were first loaded onto the C18 trap column and then separated by reverse-phase liquid chromatography (RPLC) with gradient elution. The flow rate after splitting was adjusted to 200 nL/min. The RP gradient was developed as follows: from 0 to 5% Buffer B (ACN/0.1%FA) for 5 min, from 5 to 35% Buffer B for 90 min, and from 35 to 80% Buffer B for 15 min. After running with 80% Buffer B for 10 min, the separation system was equilibrated by Buffer A (H_2O/0.1% FA) for 15 min.

For online 2D-LC–MS/MS analysis, the C18 trap column was replaced by a 7 cm phosphate monolithic trap column (200 μm i.d.). The peptide sample was first loaded onto the monolithic trap; then, a series stepwise elution with salt concentrations of 100, 200, 300, 400, 500 and 1,000 mM NH_4Ac was used to gradually elute peptides from the phosphate monolithic column onto the C18 analytical column (75 μm i.d.).[20] After the whole system was re-equilibrated for 10 min with Buffer A, the binary gradient elution previously described was applied to separate peptides prior to MS detection in each cycle.

A spray voltage of 2.2 kV was applied between the spray tip and the MS interface. The temperature of the ion transfer capillary was 250 °C. The MS/MS spectra were acquired in the data-dependent collision-induced dissociation (CID) mode. The normalized collision energy was set at 35.0%, and the activation time was 10 ms. The mass resolution was set at 60 000 for full MS. The dynamic exclusion was set as follows: repeat count, 1; duration, 30 s; exclusion list size, 500; exclusion duration, 90 s. Survey full-scan MS was acquired from m/z 400 to 2000. The 20 most intense ions were selected for MS/MS, and the minimum signal required was set at 500. System control and data collection were carried out by Xcalibur software version 2.1.

The Triple-TOF 5600 system was equipped with a NanoACQUITY UPLC (Waters, USA) for separation, including a 3 cm C18 capillary trap column (200 μm i.d.), a 20 cm C18 capillary analysis column (75 μm i.d.), and a PicoTip Emitter (New Objective, USA). The flow rate was set at 0.35 μL/min, and the RP gradient was as follows: from 0 to 5% Buffer B (98% ACN/2% H_2O/0.1% FA) for 5 min, from 5 to 25% Buffer B for 60 min, and from 35 to 70% Buffer B for 15 min. After running with 70% Buffer B for 10 min, the separation system was equilibrated by Buffer A (98% H_2O/2% ACN/0.1% FA) for 10 min. For off-line 2D-LC–MS/MS analysis, the deglycosylated peptides were first loaded onto a 7 cm phosphate monolithic trap column and then gradually eluted with 100, 200, 300, 400, 500, and 1000 mM NH_4Ac into six fractions. Finally, the fractionated samples were analyzed with the above RP LC–MS/MS system.

The Triple-TOF 5600 MS was equipped with a Digital PicoView ESI source (New Objective, USA). A spray voltage of 2.3 kV was applied between the spray tip and the MS interface. The instrument was operated in information-dependent acquisition (IDA) mode, with the top 40 precursors (charge state +2 to +5, >50 counts) in each full MS scan (800 ms, scan range m/z 350–1250) subjected to MS/MS analysis (minimum accumulation time 100 ms, scan range m/z 100–1500). Ion exclusion time was set to 30 s.

Data Analysis

All RAW files collected by Xcalibur 2.1 were searched with MaxQuant (Andromeda) version 1.1.1.36, and all the WIFF files collected by Analyst 1.5.1 were searched with ProteinPilot version 4.1.46 (AB SCIEX, USA). The IPI human 3.80 (86 719 entries) was used for both searches. For MaxQuant searching, the parameters were set as follows: enzymes, trypsin, two missed cleavages (or trypsin and Glu-C, five missed cleavages and chymotrypsin, five missed cleavages); static modification, Cys carboxamidomethylation; variable modification, Met oxidation and Asn and Gln deamidation; mass tolerance, 20 ppm for parent ions and 0.5 Da for fragment ions. For ProteinPilot searching, no variable modification and mass tolerance were set because the search engine includes all variable modifications such as Asn and Gln deamidation. The digestion was set with different enzyme specificity (trypsin/trypsin ad Glu-C/chymotrypsin), and the search effort was set as "Thorough ID" mode. The FDR for peptide identifications of both searches was controlled to <1% by adjusting the peptide score (for MaxQuant searched data) or peptide confidence (for ProteinPilot searched data). The sequence assignment was also manually evaluated to ensure the accuracy.

Sequence Motif Analysis

For the de novo derivation of sequence motifs, the Motif-X software[22] was applied to 4783 aligned N-glycosylation sites with their surrounding seven amino acids to both termini. A minimum occurrence of 20 matches was set for consensus

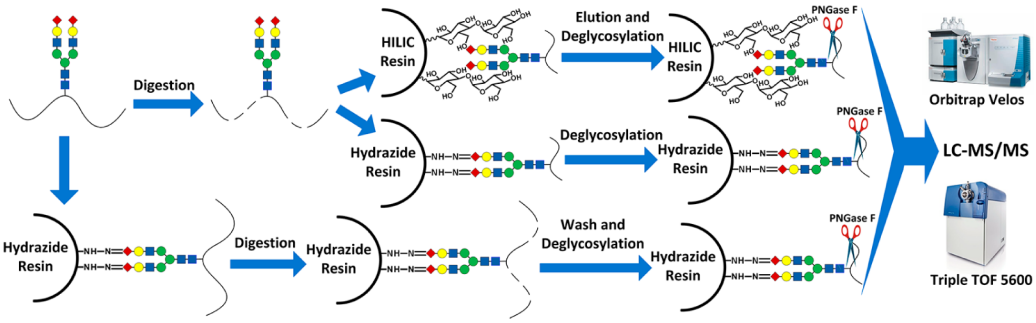


Figure 1. Workflow for comprehensive mapping of protein N-glycosylation in human liver. Protein extract of human liver was divided into three parts, one of which was enriched by HILIC strategy and the other two parts enriched by hydrazide chemistry for N-glycoproteins and N-glycopeptides. For HILIC strategy, protein was first digested and then dissolved in 80% acetonitrile before flowing through HILIC beads for enrichment. The captured N-glycopeptides were eluted by 100 mM NH_4HCO_3 aqueous buffer and deglycosylated by PNGase F. For hydrazide chemistry enrichment of N-glycopeptides, proteins were digested and then treated by $NaIO_4$ to oxidize the glycans of N-glycopeptides, and the captured N-glycopeptides were released by PNGase F. For hydrazide chemistry enrichment of N-glycoproteins, proteins were first treated by $NaIO_4$ to oxidize the glycans and then incubated with hydrazide beads. The captured N-glycoproteins were digested with proteases and then released by PNGase F. Finally, all deglycosylated peptides of the above three strategies were collected and analyzed by LC−MS/MS with two different types of mass spectrometers.

sequence identification. WebLogo[23] was used to create relative frequency plots.

Secretory Proteins Prediction

Signal Peptide Predictor (SignalP 4.1 Server, http://www.cbs.dtu.dk/services/SignalP/) was used to discriminating signal peptides from transmembrane regions.[24] SignalP used amino acid sequences to predict the existence and location of signal peptide cleavage sites. The likelihood that a protein was a signaling peptide was determined by SignalP using purely neural networks to detect signal peptides in protein sequences.

GoMiner Classifications

The cellular component, molecular function, and biological process of identified human liver N-glycosylated proteins were classified by GoMiner[25] with the new database (version_go_201202).

Biomarker Filter Analysis and Pathways Analysis

The biomarker filter analysis of the identified N-glycoproteins and the related pathway analysis were performed by Ingenuity Systems Pathway Analysis software (IPA version 8.8, Ingenuity Systems).

RESULTS AND DISCUSSION

Large-Scale Characterization of Protein N-glycosylation Sites in Human Liver

We established a workflow by the combination of click maltose−HILIC and hydrazide chemistry for the enrichment of N-glycoproteins or N-glycopeptides (Figure 1). Human liver was homogenized in 4% SDS-containing buffer, which could greatly improve the extraction efficiency of membrane proteins. For HILIC enrichment, the digestion of proteins was followed with the filter-aided sample preparation (FASP) protocol, and the peptides were enriched by using centrifugation-assisted click maltose−HILIC tips.[14,26] The hydrazide chemistry was usually carried out at protein level.[6,27] At peptide level, proteins were digested into peptides that could improve the detection of glycosylation sites buried inside proteins.[28] Therefore, the hydrazide chemistry was carried out at both protein and peptide levels in our work. In brief, at protein level, N-glycoproteins in liver extract were directly captured by hydrazide beads, and the captured N-glycopeptides were released by proteolytic digestion and PNGase F deglycosylation.[6,27] At peptide level, the hydrazide chemistry was applied to capture N-glycopeptides after proteolytic digestion of protein sample following the procedures reported previously.[6] The enrichment specificity was 60−80% for HILIC enrichment at peptide level and 85−90% for hydrazide chemistry at both protein and peptide levels in our work. Furthermore, multienzyme protein digestion by using trypsin, trypsin and Glu-C, and chymotrypsin was adopted. These three enzyme groups were complementary and yielded a 35% increase in the identification of N-glycosylation sites compared with trypsin digestion alone (Figure 2A). 62.8% of the N-glycosylation sites were identified by more than one type of enzyme group, which feasibly increases the identification confidence of N-glycosylation sites.

It is reported that during the sample preparation including proteolytic digestion, enrichment, and deglycosylation, chemical deamidation could occur to the Asn residues that do not

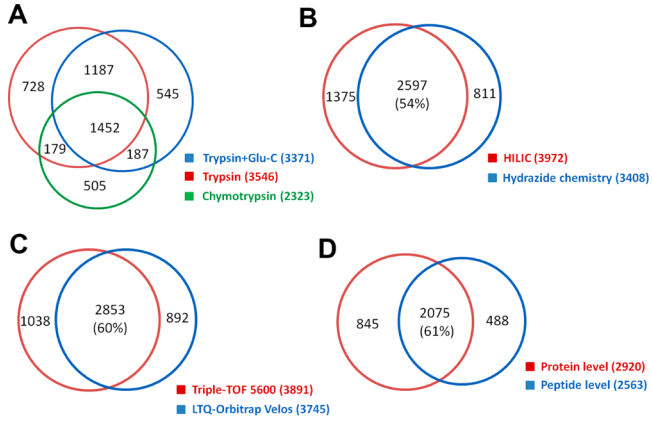


Figure 2. Venn diagrams: (A) Overlap of the identification of the N-glycosylation sites by using multienzyme digestion. (B) Overlap of the identification of the N-glycosylation sites by using two different enrichment methods (HILIC and hydrazide chemistry). (C) Overlap of the identification of the N-glycosylation sites by using two different types of mass spectrometers (LTQ-Orbitrap Velos and Triple-TOF 5600). (D) Overlap of the identification of the N-glycosylation sites enriched by hydrazide chemistry at peptide or protein level.

dx.doi.org/10.1021/pr401200h | J. Proteome Res. 2014, 13, 1713−1721

contain any modifications.[29,30] Therefore, it is necessary to evaluate the frequency of chemical deamidation in N-glycoproteomic experiments. In brief, the liver proteins after digestion and enrichment were divided into two parts, and then one part was deglycosylated using PNGase F at 37 °C overnight, while the other part was incubated at 37 °C overnight without adding any PNGase F. Finally, the two samples were both analyzed with LC−MS/MS and the results were compared (Table S1 in the Supporting Information). Interestingly, very few peptides with chemical deamidation were identified, and only 9, 22, and 6 were matched with N-!P-[S|T|C] sequence motif in peptide-level hydrazide chemistry, protein-level hydrazide chemistry, and HILIC experiments, respectively. This is mainly because the enrichment specificity was very high (85−90% for hydrazide chemistry and 60−80% for HILIC), few nonspecifically captured peptides containing Asn residues were left in our samples, and those matched with N-!P-[S|T|C] motif were even fewer. Meanwhile, we identified 1100, 1500, and 1019 N-glycopeptides in the comparing PNGase F-treated samples using peptide-level hydrazide chemistry, protein-level hydrazide chemistry, and HILIC methods for the enrichment, respectively. Above all, the chemical deamidation accounts for only 1 to 2% of the whole identified N-glycopeptides in our experiments.

Finally, the enriched N-glycopeptides were separated by reversed-phase 1-D liquid chromatography (1D-LC) or strong cation exchange-reversed phase 2-D liquid chromatography (SCX-RP 2D-LC) and analyzed by both LTQ-Orbitrap Velos MS and Triple-TOF 5600 MS. Altogether, we identified 14 480 N-glycopeptides with high confidence corresponding to 2210 N-glycoproteins and 4783 N-glycosylation sites from the human liver tissue (Table S2 in the Supporting Information). To the best of our knowledge, this is the largest human N-glycoproteome data set to date. According to the SwissProt database,[31] we covered ∼64% of the known human N-glycosylation sites and improved the human N-glycosylation sites by 158%. The identification of N-glycosylation sites was also highly reproducible, and the sites overlap was 60−70% between two replicates analyses. Similar numbers of N-glycosylation sites were identified by Triple-TOF 5600 MS (3891) and LTQ-Orbitrap Velos MS (3745), and the overlap between them was as high as 60% (2853). Moreover, we identified 3972 and 3408 N-glycosylation sites by using HILIC and hydrazide chemistry, respectively. The overlap between two enrichment methods was 54% (2597). Because N-glycopeptides and N-glycoproteins were captured by their attached N-glycans, there is no significant difference in the sequence motifs between HILIC and hydrazide chemistry. The combination of HILIC and hydrazide chemistry methods is helpful to increase the N-glycoproteome coverage due to the enrichment complementarities of these two methods. All of the above results demonstrated that our methods were robust and the data set was highly confident (Table S2 in the Supporting Information and Figure 2A−C).

Analysis of Sequence Motifs and N-glycosylation Sites Distribution

N-!P-[S|T|C] is the consensus N-glycosylation sequence motif.[9] 4783 unique N-glycosylation sites matched with this motif were identified in our case, corresponding to 4667 nonredundant sequence windows. Proline is extremely underrepresented in the first position relative to the modified asparagines (frequency = 0%), as expected. 4521 sequence windows have serine or threonine at the second position to the C terminus of modified asparagines, while threonine occurs more frequently (1.3-fold) than serine at the second position (2561 versus 1960). 146 sequence windows were found to be heavily enriched for cysteine in place of S/T, 51 of which were identified in both methods of N-glycopeptides enrichment (Figure 3A).

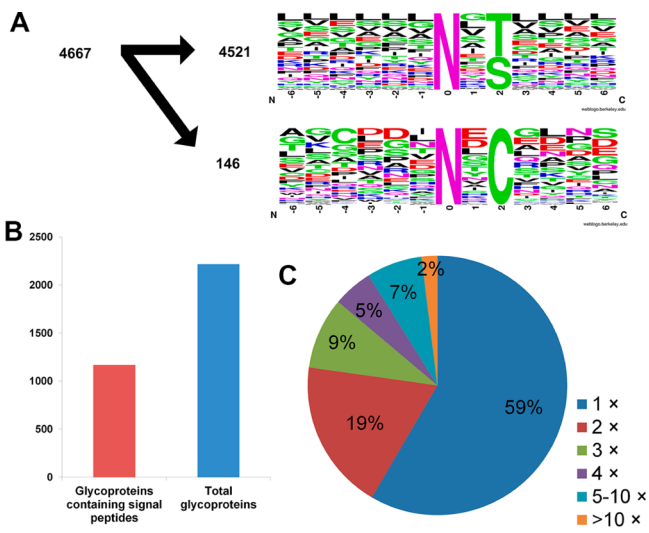


Figure 3. Analyses of sequence motifs, secondary structures, and N-glycosylation sites distribution. (A) Distribution of the 4667 nonredundant sequence windows matching with N-!P-[S|T] and N-!P−C motifs. (B) Proportion of proteins containing signal peptides. (C) Distribution of single and multiple N-glycosylation.

The signal peptide of the corresponding 2210 N-glycoproteins was predicted by Signal P4.1.[24] About 52% of the N-glycoproteins were predicted with signal peptide (Figure 3B). The number of identified N-glycosylation sites in each protein was also investigated (Figure 3C). Among 2210 N-glycoproteins, 59% were detected with a single N-glycosylation site, 19 and 9% were with two and three N-glycosylation sites, respectively, and the average degree of N-glycosylation was 2.2. Notably, there was a group of 166 proteins that contained five or more N-glycosylation sites and 39 with at least 10 N-glycosylation sites. Prolow-density lipoprotein receptor-related protein 1 (LRP1) was most heavily N-glycosylated with 60 N-glycosylation sites. Besides, fibrillin-1 and stabilin-1 were also heavily glycosylated with 29 and 27 N-glycosylation sites, respectively.

Cellular Component, Molecular Function, and Biological Process

An overview of the cellular component, molecular function, and biological process of the identified N-glycoproteins in human liver was obtained by applying GoMiner[25] with the new database (version_go_201202). Meanwhile, the Liverbase version 2.0 updated in March 2012 (http://liverbase.hupo.org.cn/index2.jsp) was also analyzed by GoMiner to get a full view of the whole liver proteome. Among the total 2210 N-glycoproteins in our human N-glycoproteome data set, 1476 N-glycoproteins (66.8%) were found to be gene ontology (GO)-annotated proteins. Most of the GO-annotated N-glycoproteins were associated with the processes of N-glycoprotein synthesis and transport, for example, membrane and extracellular regions, which are the transport destinations of glycoproteins; endoplasmic reticulum and Golgi apparatus, which are the

dx.doi.org/10.1021/pr401200h | J. Proteome Res. 2014, 13, 1713−1721

major organelles of glycoprotein synthesis; and vesicle which is the organelle responsible for transport between intracellular and extracellular environment. Compared with the subcellular distribution of the proteome data set (Liverbase 2.0), the proportion of proteins in our data set related to subcellular locations previously mentioned was dramatically increased (Figure 4A). Moreover, the characteristic molecular functions

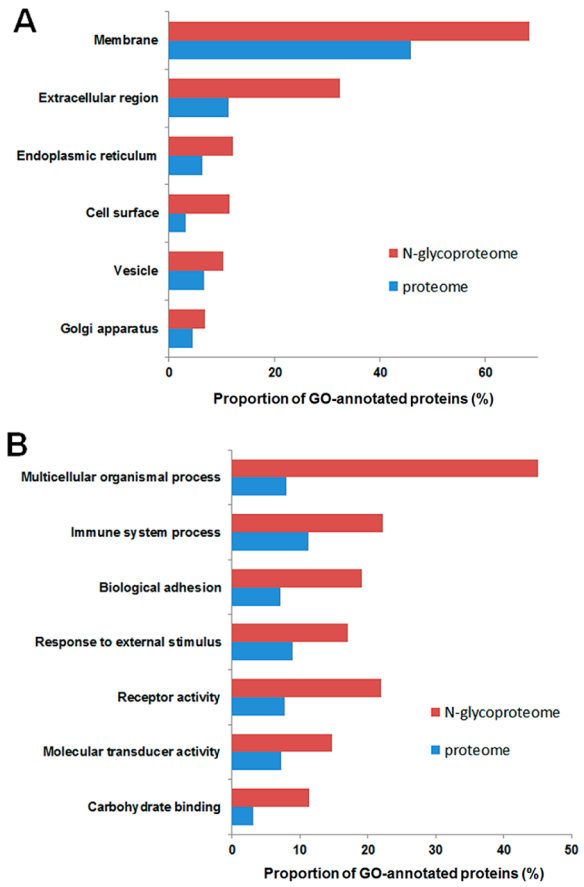


Figure 4. Cellular components (A) and molecular functions and biological processes (B) that were remarkably overexpressed in our N-glycoproteome data set compared with the human Liverbase 2.0, which is the newly updated liver proteome data set, according to the results of GoMiner analysis.

of "carbohydrate binding", "receptor activity", and "molecular transducer activity" for N-glycoproteins and the featured biological processes participated with membrane proteins and secreted proteins of "multicellular organismal process", "immune system process", "biological adhesion", and "response to external stimulus" were all significantly enriched in our N-glycoproteome data set relative to the Liverbase 2.0, which is consistent with the results of mouse tissues obtained by Zielinska et al.[9] (Figure 4B).

Among the total of 2210 N-glycoproteins, 353 are related to different types of human diseases in biomarker filter analysis by Ingenuity Pathway Analysis (IPA), such as cancer, cardiovascular disease, and auditory disease. And 269 are involved in human liver diseases, most of which are liver cancer (Table S3 in the Supporting Information). Therefore, our N-glycoproteins data set covers a considerable number of marker candidates of human liver diseases, which may be helpful for understanding the biological functions of these proteins. For example, E-cadherin (CDH1) is a key trans-membrane component of adherens junctions in cancer progression, and appeared as an independent predictor of recurrence of disease (ROD) of hepatocellular carcinoma (HCC).[32,33] Our results demonstrate that the E-cadherin N577 (DFEHVKNSTYTAL) and N656 (THGASANWTIQYN) locating outside the membrane can be N-glycosylated in human liver, which may be important regulators in CDH1 interactions, such as with β-catenin and serine/threonine protein phosphatase 2A (PP2A).[34,35]

To improve the coverage of human liver N-glycoproteome, we applied the new click maltose−HILIC and the improved hydrazide chemistry in this work. The click maltose−HILIC was previously demonstrated to have much better enrichment efficiency and specificity compared with commercial available HILIC materials.[36] By using the new click maltose−HILIC and Triple-TOF 5600 MS, we were able to identify more than 800 N-glycosylation sites in a single 1D-LC−MS/MS run with only 90 min gradient and 40 μg protein sample. In a 2D SCX-RP-LC−MS/MS run with six salt-gradient steps, we could identify over 1600 N-glycosylation sites (Table S4 in the Supporting Information). For the hydrazide chemistry enrichment of both glycosylated proteins and peptides, an optimized oxidation buffer containing 2% SDS and 2% triton was adopted to improve the solubility of membrane proteins. In the preliminary experiment, the glycopeptides of mouse liver that were digested with trypsin were captured onto the hydrazide resins. It was observed that when adding detergent in the oxidation buffer, the identification of N-glycosylation sites in a single run of 1D LC−MS/MS (LTQ-Orbitrap Velos MS) analysis was improved by more than 55% (Table S5 in the Supporting Information). Therefore, the appendix of 1% SDS and 1% triton in the urea solution during digestion of the glycosylated proteins captured on the hydrazide resins was applied for the same reason. Additionally, the enrichment of both glycosylated proteins and peptides could improve the number of N-glycosylation by 34% compared with enriching glycosylated peptides alone (Figure 2D). Finally, the combination of both maltose−HILIC and the improved hydrazide chemistry-based enrichment methods increased the identification of N-glycosylation by 40% compared with using one enrichment method alone (Figure 2B).

There is little evidence of mouse model and human cell lines that can faithfully reflect all of the physiological and pathological phenomena of a specific human disease. The defects on these model systems may mislead and confuse the investigation on disease mechanism and therapy. Therefore, it is essential to validate the discoveries on human after model system study. Although over tens of thousands of N-glycosylation sites are feasibly identified, the comprehensive mapping of N-glycosylation sites of human is still lacking. In this study, the N-glycoproteome of human liver was analyzed in-depth by a combination of HILIC and hydrazide chemistry enrichment as well as different MS detection. Finally, 4783 unique N-glycosylation sites matched with N-!P-[S|T|C] motif were reliably identified from 14 480 N-glycopeptides, and 3414 N-glycosylation sites corresponding to 1373 N-glycoproteins were identified for the first time, which greatly expands our knowledge of human liver N-glycosylated substrates and exact N-glycosylation sites. For example, the gene of human leukocyte antigen class II complex (HLA-DQ) is crucial to immune-mediated diseases, such as liver diseases and cancers, and was recently reported to be significantly associated with HCC risk.[37] In our study, N-glycosylation was detected on

N104 and 144 of HLA-DQA1 and N51 of HLA-DQB1. Furthermore, N-glycosylation was also observed on other HLA antigens, such as HLA-A/B/C/E/F, HLA-DM, HLA-DP, and HLA-DR (Table S6in the Supporting Information). As another example, G-protein-coupled receptors (GPCRs) include a large protein family of receptors and are also the target of ∼40% of all modern medicinal drugs.[38] N-glycosylation plays crucial roles in GPCR-signaling based cellular regulation.[39] 65 N-glycosylation sites on 29 GPCRs were identified in our study (Figure 5 and Table S7 in the Supporting Information). Such detailed N-glycosylation information may be useful for in-depth study of the GPCRs biological functions.

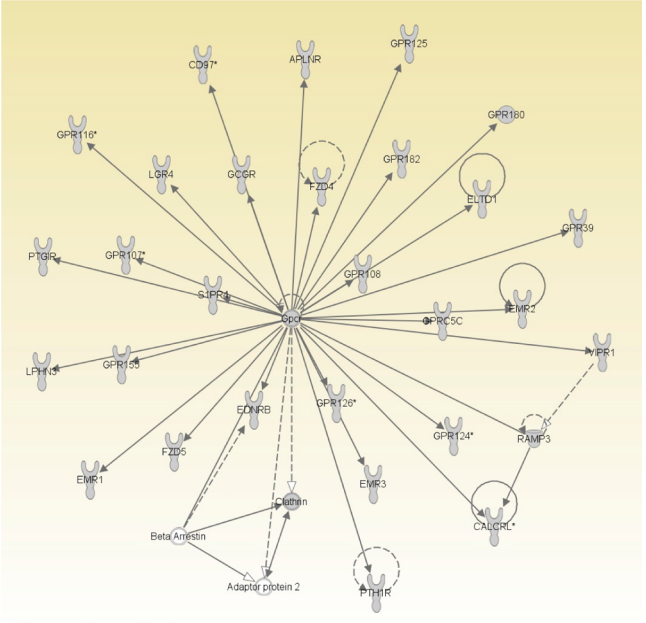


Figure 5. N-glycosylated G protein coupled receptors. 65 N-glycosylation sites on 29 G protein coupled receptors (gray) were identified in this study.

N-Glycosylation of Proteins Involved into Hepatic Stellate Cell Activation

Hepatic stellate cells (HSCs) are liver-specific pericytes that stores vitamin A and are considered to be a key point of hepatic fibrosis. Usually, HSCs are in the quiescent state within normal liver. However, HSCs are switched to the activated state in response to chronic liver injury. The activated HSCs trans-differentiate to myofibroblasts and are the source of secreting extracellular matrixes (ECMs), which will increase the stromal stiffness and induce fibrogenesis. The liver fibrogenesis directly supports hepatic tumorigenesis and is strongly associated with HCC.[40,41] 50 proteins involved into HSC activation (hepatic fibrosis) were N-glycosylated in our study, corresponding to 173 N-glycosylation sites (Table S8 in the Supporting Information). Many cytokines are secreted by HSC in human liver, and these cytokines and their membrane receptors are essential to cell−cell interactions in both normal and injured liver.[42] It was observed that most of the protein receptors on cellular membranes of the early and activated HSCs are glycosylated (Figure S1 in the Supporting Information), such as platelet-derived growth factor receptor (PDGFR), leptin receptor (LEPR), tumor necrosis factor receptor (TNFR), transforming growth factor beta receptor (TGFβR), epidermal growth factor receptor (EGFR), and interleukin-4 receptor (IL4R). The TGFβ- and PDGF-dependent tumor−stroma cross-talk was reported to enhance the growth and migration of malignant hepatocytes.[43] PDGF and its receptor PDGFR were up-regulated during liver injury, and the overexpression of PDGF leads to both fibrosis and HCC in mouse.[44] Activated PDGFR recruits the signaling molecule Ras and initiates ERK/MAP kinase pathway, which are both related to cancer progression.[45] TGFβR activates the Smad signaling, which plays different roles in the progression of cellular activation.[46] Therefore, our study uncovers the wide existence of N-glycosylation in both quiescent and activated HSC and gives the exact N-glycosylation sites on many growth factors, cytokines, and membrane receptors. Because few efforts have been made on this aspect, further elucidating the biological function of N-glycosylation in HSC activation might be important for the treatment of fibrogenesis and cirrhosis.

N-Glycoproteins in Acute Phase Response of Human Liver

The acute phase response is a rapid inflammatory response triggered by infection, tissue injury, trauma or surgery, neoplastic growth, or immunological disorders. About 15% of human cancers are associated with chronic infections and inflammation.[47] Liver is an important constituent of the human immune system and also the major source of acute phase response proteins. Interleukin-1 (IL-1) and interleukin-6 (IL-6) type cytokines are leading regulators to initiate the cascade of acute-phase protein expression,[48] and the NF-κB and STAT3 signaling pathway control the expression of a series of growth factors and cytokines within hepatocytes in responding to liver inflammatory responses, which are critical for compensatory liver regeneration and HCC development.[49] We feasibly detected 50 N-glycoproteins corresponding to 186 N-glycosylation sites within the acute phase response signaling (Table S9 in the Supporting Information). Obviously, more than half of the regulating proteins involved in initiating the expression cascade of acute phase response are N-glycosylated, such as tumor necrosis factor receptor (TNFR), IL-1, IL-1 receptor (IL-1R), IL-6 receptor (IL-6R), IL-6 receptor subunit beta (GP130), and oncostatin-M-specific receptor (OSMR). Furthermore, N-glycosylation can be found on most of the expressed acute phase response proteins. For example, the NF-κB-dependent cytokines, such as serum amyloid A (SAA), complement factor B (CFB), complement C3 (C3), fibronectin 1 (FN1), orosomucoid (ORM), angiotensinogen (AGT), and superoxide dismutase 2 (SOD2), and STAT3-dependent cytokines, such as serum amyloid P-component (APCS), serpin peptidase inhibitor clade A (SERPINA), fibrinogen gamma chain (FGG), fibrinogen beta chain (FGB), fibrinogen alpha chain (FGA), alpha-2-macroglobulin (A2M), and lipopolysaccharide-binding protein (LBP), are all detected in N-glycoproteome analysis with exact N-glycosylation sites. The other proteins relating to NF-IL6- and TCF-dependent acute phase response are also widely N-glycosylated (Figure S2 in the Supporting Information). Therefore, protein N-glycosylation is also widely involved in acute phase response of human liver, and the related biological functions merit further investigation.

■ CONCLUSIONS

In this work, in total 14 498 N-glycosylated peptides matched with N-!P-[S|T|C] sequence motif were confidently identified from human liver, corresponding to 2210 N-glycosylated proteins and 4783 N-glycosylation sites. It is clear that the

dx.doi.org/10.1021/pr401200h | J. Proteome Res. 2014, 13, 1713−1721

comprehensive mapping of protein N-glycosylation sites in human liver generated a valuable public data set for the study of various molecular and cellular processes, which are highly related to different types of human diseases. We hope this data set can also provide some inspirations to the community for detailed studies of functional biology.

■ ASSOCIATED CONTENT

Supporting Information

Signaling events in the early and activated HSC. Signaling events in acute phase response of human liver. Investigation results of chemical deamidation in the click maltose HILIC and hydrazide chemistry-enriched human liver samples. Detailed information about the N-glycosylated proteins and N-glycosylation sites from human liver identified by Triple-TOF 5600 and LTQ-Orbitrap Velos. Detailed information about identified N-glycosylated proteins and N-glycosylation sites involved in human liver diseases. N-glycosylation sites identified from human liver by a single 90 min run of 1D-LC−MS/MS and a 2D-SCX-RP LC−MS/MS run with seven salt-gradient steps with the enrichment of HILIC. Detailed information of N-glycosylation sites identified from mouse liver with oxidation buffer and optimized oxidation buffers. Detailed information about identified N-glycosylation sites on HLA antigens. Detailed information about identified N-glycosylation sites on G protein coupled receptors. Detailed information about identified N-glycosylated proteins and N-glycosylation sites involved in HSC activation (hepatic fibrosis). Detailed information about identified N-glycosylated proteins and N-glycosylation sites involved in acute phase response signaling. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*F.W.: E-mail: wangfj@dicp.ac.cn. Fax: (+) 86-411-84379620. Phone: (+) 86-411-84379610.
*H.Z.: E-mail: hanfazou@dicp.ac.cn.

Author Contributions

§J.Z. and Z.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We greatly appreciate Dr. Daniel Figeys for the help of Ingenuity Pathway Analysis (IPA) and Dr. Xinmiao Liang and Dr. Zhimou Guo for providing the HILIC materials as a gift. H.Z. acknowledges the China State Key Basic Research Program Grant (2013CB-911202, 2012CB910101), the financial support from the NSFC (21321064, 81161120540, and 21235006), the National Key Special Program on Infection diseases (2012ZX10002009-011), and the Analytical Method Innovation Program of MOST (2012IM030900). F.W. acknowledges the financial support from NSFC (21305139) and the "Hundred Talent Young Scientist Program" by DICP.

■ REFERENCES

(1) Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *BBA, Biochim. Biophys. Acta, Gen. Subj.* **1999**, *1473* (1), 4−8.

(2) Ott, C. M.; Lingappa, V. R. Integral membrane protein biosynthesis: why topology is hard to predict. *J. Cell Sci.* **2002**, *115* (10), 2003−2009.

(3) Roth, J. Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **2002**, *102* (2), 285−303.

(4) Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002**, *12* (4), 43R−56R.

(5) Glinsky, G. V. Antigen presentation, aberrant glycosylation and tumor progression. *Crit. Rev. Oncol./Hematol.* **1994**, *17* (1), 27−51.

(6) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21* (6), 660−666.

(7) Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **2005**, *4* (6), 2070−2080.

(8) Zhang, H.; Yi, E. C.; Li, X.-j.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G.; Smith, R. D.; Kemp, C. J.; Aebersold, R. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **2005**, *4* (2), 144−155.

(9) Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **2010**, *141* (5), 897−907.

(10) Zielinska, D. F.; Gnad, F.; Schropp, K.; Wi; Mann, M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* **2012**, *46* (4), 542−548.

(11) Kaji, H.; Shikanai, T.; Sasaki-Sawa, A.; Wen, H.; Fujita, M.; Suzuki, Y.; Sugahara, D.; Sawaki, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Takahashi, N.; Isobe, T.; Narimatsu, H. Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.* **2012**, *11* (9), 4553−4566.

(12) Chen, R.; Jiang, X. N.; Sun, D. G.; Han, G. H.; Wang, F. J.; Ye, M. L.; Wang, L. M.; Zou, H. F. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J. Proteome Res.* **2009**, *8* (2), 651−661.

(13) Parker, B. L.; Palmisano, G.; Edwards, A. V. G.; White, M. Y.; Engholm-Keller, K.; Lee, A.; Scott, N. E.; Kolarich, D.; Hambly, B. D.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J., Quantitative n-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment. *Mol. Cell. Proteomics* **2011**, *10*, (8).

(14) Zhu, J.; Wang, F.; Chen, R.; Cheng, K.; Xu, B.; Guo, Z.; Liang, X.; Ye, M.; Zou, H. Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. *Anal. Chem.* **2012**, *84* (11), 5146−5153.

(15) Jemal, A.; Bray, F.; Center, M. M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *Ca-Cancer J. Clin.* **2011**, *61* (2), 69−90.

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

(17) He, F. Human liver proteome project: plan, progress, and perspectives. *Mol. Cell. Proteomics* **2005**, *4* (12), 1841−1848.

(18) Jiang, Y.; Ying, W.; Wu, S.; Chen, M.; Guan, W.; Yang, D.; Song, Y.; Liu, X.; Li, J.; Hao, Y.; Sun, A.; Geng, C.; Li, H.; Mi, W.; Zhang, Y.; Zhang, J.; Chen, X.; Li, L.; Gong, Y.; Li, T.; Ma, J.; Li, D.; Yuan, X.; Zhang, X.; Xue, X.; Zhu, Y.; Qian, X.; He, F.; Zhong, F.; Shen, H.; Lin, C.; Lu, H.; Liu, X.; Wei, L.; Cao, J.; Yun, D.; Zhang, J.; Gao, M.; Fan, H.; Zhang, Y.; Cheng, G.; Yu, Y.; Xie, L.; Wang, H.; Zhang, X.; Yang, P.; Shi, L.; Tong, W.; Li, X.; Wang, Y.; Liu, S.; Sheng, Q.; Zeng, R.; Sun, Y.; Xu, Y.; Cai, J.; He, P.; Gao, H.; Zhao, X.; Tan, Y.; Yan, H.; Yang, Y.; Wang, H.; Huang, J.; Han, Z.; He, Q.; Chen, P.; Liang, S.;

Zhao, M.; Mao, X.; Wei, L.; Yu, H.; Cao, Z.; Li, Y.; Dai, W.; Jiang, H.; Wang, D.; Zheng, J.; Gao, X.; Tang, Y.; Li, X.; Cheng, J.; Liu, Y.; Zhang, X.; Wang, X.; Jia, J.; An, D.; Wang, Z.; Li, Q.; Cui, T.; Chinese Human Liver, P. First insight into the human liver proteome from PROTEOMESKY-LIVERHu 1.0, a publicly available database. *J. Proteome Res.* **2010**, *9* (1), 79−94.

(19) Sun, A.; Jiang, Y.; Wang, X.; Liu, Q.; Zhong, F.; He, Q.; Guan, W.; Li, H.; Sun, Y.; Shi, L.; Yu, H.; Yang, D.; Xu, Y.; Song, Y.; Tong, W.; Li, D.; Lin, C.; Hao, Y.; Geng, C.; Yun, D.; Zhang, X.; Yuan, X.; Chen, P.; Zhu, Y.; Li, Y.; Liang, S.; Zhao, X.; Liu, S.; He, F. Liverbase: A Comprehensive View of Human Liver Biology. *J. Proteome Res.* **2009**, *9* (1), 50−58.

(20) Wang, F. J.; Dong, J.; Jiang, X. G.; Ye, M. L.; Zou, H. F. Capillary trap column with strong cation-exchange monolith for automated shotgun proteome analysis. *Anal. Chem.* **2007**, *79* (17), 6599−6606.

(21) Wang, F. J.; Chen, R.; Zhu, J.; Sun, D. G.; Song, C. X.; Wu, Y. F.; Ye, M. L.; Wang, L. M.; Zou, H. F. A fully automated system with online sample loading, isotope dimethyl labeling and multidimensional separation for high-throughput quantitative proteome analysis. *Anal. Chem.* **2010**, *82* (7), 3007−3015.

(22) Schwartz, D.; Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **2005**, *23* (11), 1391−1398.

(23) Thomas, P. D.; Kejariwal, A.; Campbell, M. J.; Mi, H.; Diemer, K.; Guo, N.; Ladunga, I.; Ulitsky-Lazareva, B.; Muruganujan, A.; Rabkin, S.; Vandergriff, J. A.; Doremieux, O. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **2003**, *31* (1), 334−341.

(24) Petersen, T. N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8* (10), 785−786.

(25) Zeeberg, B. R.; Feng, W. M.; Wang, G.; Wang, M. D.; Fojo, A. T.; Sunshine, M.; Narasimhan, S.; Kane, D. W.; Reinhold, W. C.; Lababidi, S.; Bussey, K. J.; Riss, J.; Barrett, J. C.; Weinstein, J. N., GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **2003**, *4*, (4).

(26) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6* (5), 359−62.

(27) Sun, B. Y.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X. W.; Lin, B. Y.; Hood, L. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol. Cell. Proteomics* **2007**, *6* (1), 141−149.

(28) Rudd, P. M.; Dwek, R. A. Glycosylation: Heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32* (1), 1−100.

(29) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K., Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation. *Mol. Cell. Proteomics* **2011**, *10*, (10).

(30) Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* **2012**, *11* (3), 1949−57.

(31) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1*, 1−5.

(32) Fiorentino, M.; Altimari, A.; Ravaioli, M.; Gruppioni, E.; Gabusi, E.; Corti, B.; Vivarelli, M.; Bringuier, P. P.; Scoazec, J. Y.; Grigioni, W. F.; D'Errico-Grigioni, A. Predictive value of biological markers for hepatocellular carcinoma patients treated with orthotopic liver transplantation. *Clin. Cancer Res.* **2004**, *10* (5), 1789−95.

(33) Huber, A. H.; Weis, W. I. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **2001**, *105* (3), 391−402.

(34) Jamal, B. T.; Nita-Lazar, M.; Gao, Z.; Amin, B.; Walker, J.; Kukuruzinska, M. A. N-glycosylation status of E-cadherin controls cytoskeletal dynamics through the organization of distinct beta-catenin- and gamma-catenin-containing AJs. *Cell Health Cytoskeleton* **2009**, *2009* (1), 67−80.

(35) Liwosz, A.; Lei, T.; Kukuruzinska, M. A. N-glycosylation affects the molecular organization and stability of E-cadherin junctions. *J. Biol. Chem.* **2006**, *281* (32), 23138−49.

(36) Yu, L.; Li, X.; Guo, Z.; Zhang, X.; Liang, X. Hydrophilic interaction chromatography based enrichment of glycopeptides by using click maltose: A matrix with high selectivity and glycosylation heterogeneity coverage. *Chem.—Eur. J.* **2009**, *15* (46), 12618−12626.

(37) Jiang, D. K.; Sun, J.; Cao, G.; Liu, Y.; Lin, D.; Gao, Y. Z.; Ren, W. H.; Long, X. D.; Zhang, H.; Ma, X. P.; Wang, Z.; Jiang, W.; Chen, T. Y.; Gao, Y.; Sun, L. D.; Long, J. R.; Huang, H. X.; Wang, D.; Yu, H.; Zhang, P.; Tang, L. S.; Peng, B.; Cai, H.; Liu, T. T.; Zhou, P.; Liu, F.; Lin, X.; Tao, S.; Wan, B.; Sai-Yin, H. X.; Qin, L. X.; Yin, J.; Liu, L.; Wu, C.; Pei, Y.; Zhou, Y. F.; Zhai, Y.; Lu, P. X.; Tan, A.; Zuo, X. B.; Fan, J.; Chang, J.; Gu, X.; Wang, N. J.; Li, Y.; Liu, Y. K.; Zhai, K.; Hu, Z.; Liu, J.; Yi, Q.; Xiang, Y.; Shi, R.; Ding, Q.; Zheng, W.; Shu, X. O.; Mo, Z.; Shugart, Y. Y.; Zhang, X. J.; Zhou, G.; Shen, H.; Zheng, S. L.; Xu, J.; Yu, L. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat. Genet.* **2012**, *45* (1), 72−5.

(38) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5* (12), 993−6.

(39) Tansky, M. F.; Pothoulakis, C.; Leeman, S. E. Functional consequences of alteration of N-linked glycosylation sites on the neurokinin 1 receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (25), 10691−10696.

(40) Friedman, S. L. Evolving challenges in hepatic fibrosis. *Nat. Rev. Gastroenterol Hepatol.* **2010**, *7* (8), 425−36.

(41) Zhang, D. Y.; Friedman, S. L. Fibrosis-dependent mechanisms of hepatocarcinogenesis. *Hepatology* **2012**, *56* (2), 769−75.

(42) Friedman, S. L. Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. *Physiol. Rev.* **2008**, *88* (1), 125−72.

(43) van Zijl, F.; Mair, M.; Csiszar, A.; Schneller, D.; Zulehner, G.; Huber, H.; Eferl, R.; Beug, H.; Dolznig, H.; Mikulits, W. Hepatic tumor-stroma crosstalk guides epithelial to mesenchymal transition at the tumor edge. *Oncogene* **2009**, *28* (45), 4022−33.

(44) Czochra, P.; Klopcic, B.; Meyer, E.; Herkel, J.; Garcia-Lazaro, J. F.; Thieringer, F.; Schirmacher, P.; Biesterfeld, S.; Galle, P. R.; Lohse, A. W.; Kanzler, S. Liver fibrosis induced by hepatic overexpression of PDGF-B in transgenic mice. *J. Hepatol.* **2006**, *45* (3), 419−28.

(45) Carloni, V.; Defranco, R. M.; Caligiuri, A.; Gentilini, A.; Sciammetta, S. C.; Baldi, E.; Lottini, B.; Gentilini, P.; Pinzani, M. Cell adhesion regulates platelet-derived growth factor-induced MAP kinase and PI-3 kinase activation in stellate cells. *Hepatology* **2002**, *36* (3), 582−91.

(46) Uemura, M.; Swenson, E. S.; Gaca, M. D.; Giordano, F. J.; Reiss, M.; Wells, R. G. Smad2 and Smad3 play different roles in rat hepatic stellate cell function and alpha-smooth muscle actin organization. *Mol. Biol. Cell* **2005**, *16* (9), 4214−24.

(47) Coussens, L. M.; Werb, Z. Inflammation and cancer. *Nature* **2002**, *420* (6917), 860−7.

(48) Bode, J. G.; Albrecht, U.; Haussinger, D.; Heinrich, P. C.; Schaper, F. Hepatic acute phase proteins–regulation by IL-6- and IL-1-type cytokines involving STAT3 and its crosstalk with NF-kappaB-dependent signaling. *Eur. J. Cell Biol.* **2012**, *91* (6−7), 496−505.

(49) He, G.; Karin, M. NF-kappaB and STAT3 - key players in liver inflammation and cancer. *Cell. Res.* **2011**, *21* (1), 159−68.