Supporting Information

A fully integrated sample preparation strategy for highly sensitive intact glycoproteomics

Lijun Yang,[#, a] Jie Liu,[#, a] Hua Li,[b] Yilian Liu,[c] An He,[c] Peiwu Huang,[c] Weina Gao,[c] Hua Cao,[a] Ruilian Xu,[*, a] Ruijun Tian[*, c]

[a] Department of Oncology, The Second Clinical Medical College, Jinan University (Shenzhen People's Hospital), Shenzhen 518020, China. The First Affiliated Hospital, Jinan University, Guangzhou 510632, China.

[b] SUSTech Core Research Facilities, Southern University of Science and Technology, Shenzhen 518055, China.

[c] Department of Chemistry and Guangdong Provincial Key Laboratory of Cell Microenvironment and Disease Research, Southern University of Science and Technology, Shenzhen 518055, China.

# These authors contribute equally to this manuscript.

Corresponding author:

* Ruilian, Xu, Tel: +86-0755-22942709, E-mail: xuruilian@126.com;

* Ruijun Tian, Tel: +86-755-88018905, E-mail: tianrj@sustech.edu.cn.

Contents of SI materials

Materials and methods

Cell culture and secretome preparation

Intact GlycoSISPROT tip fabrication

Intact glycopeptide enrichment using Intact GlycoSISPROT

Intact glycopeptide enrichment using traditional method

LC-MS/MS Analysis

Data Analysis

Results and discussion

Gene Ontology annotation for molecular function of glycoproteins identified by Intact GlycoSISPROT

Supporting Figures

Fig. S1 Optimization of the Intact GlycoSISPROT for intact glycopeptide analysis starting from 1 μg IgG.

Fig. S2 Glycosylation analysis of 1 μg IgG by Intact GlycoSISPROT.

Fig. S3 The MS/MS spectra of the intact glycopeptides from IgG.

Fig. S4 Glycosylation analysis of fetuin by Intact GlycoSISPROT.

Fig. S5 Sensitivity of the Intact GlycoSISPROT.

Fig. S6 Venn diagram of site-specific glycoforms identified from 1 μg of secreted proteins.

Fig. S7 Glycosylation heterogeneity of TIMP1.

Fig. S8 Gene Ontology annotation for molecular function of glycoproteins identified from 25 μg of secreted proteins by the Intact GlycoSISPROT.

Reference

Materials and methods

Cell culture and secretome preparation

The KP4 pancreatic cancer cells obtained from ATCC were cultured in RPMI 1640 (Corning cellgro) supplemented with 10% (v/v) fetal bovine serum (PAN Biotech), 1% penicillin and streptomycin (Corning cellgro) in a 37 °C incubator supplied with 5% CO_2. The cells with around 80% confluence were washed with RPMI 1640 for three times and then cultured in serum-free medium for 24-36 h. Conditioned medium was centrifugated twice to remove the cell debris, further cleaned by passing through a 0.45 μm filter (Millipore) and concentrated by centrifugating through a centrifugal filter (Millipore) with 3 kDa molecular mass cut-off at 4,600 *g* for 45 min at 4 °C. The sample was exchanged with PBS for three times in the same centrifugal filter and quantified by BCA assay (Sigma).

Intact GlycoSISPROT tip fabrication

The Intact GlycoSISPROT tip is assembled by packing SCX/SAX mixed beads (Applied Biosystems) to bind protein sample, C18 disk (3M) for peptide desalting and ZIC-HILIC beads (Merck, Product no. 1.50458.0001) for intact glycopeptide enrichment (**Figure 1a**). Zwitterionic-HILIC (ZIC-HILIC) is a popular type of HILIC stationary phase for unbiased and efficient enrichment of intact glycopeptides [1]. Firstly, a small piece of cotton wool was packed into a 10 μL pipet tip. And then 0.5 mg ZIC-HILIC beads, 1 plug C18 disk and 0.4 mg SCX/SAX mixed beads (1:1, w/w) were loaded sequentially into the tip for processing 1 μg protein sample. The amount of SCX/SAX mixed beads, C18 disk and ZIC-HILIC beads were adjusted according to the protein amount. According to our experience, 0.5 mg ZIC-HILIC beads is used for 1 μg digested peptides, and one plug of C18 disk is used for 5 μg of digested peptides. 1 mg POROS SCX/SAX mixed beads are used for 20 μg proteins.

Intact glycopeptide enrichment using Intact GlycoSISPROT

Firstly, Intact GlycoSISPROT tip was washed with 60 μL methanol and equilibrated with 20 μL of 100 mM potassium citrate buffer (CAK), pH 3 and 20 μL of 10 mM CAK, pH 3 by centrifugation sequentially. IgG (sigma), fetuin (New England Biolabs), and KP4 cell secretome was acidified to pH 2-3 and loaded by centrifugation for 5 min. The loading procedure was repeated once. The tip was then washed with 20 μL 20% (v/v) ACN in 8 mM CAK and 40 μL ACN sequentially. Then 20 μL 10 mM TCEP in 9 mM CAK, pH 3 was added and incubated for 15 min at room temperature. After the reduction step, the tip was washed with 20 μL 50 mM Tris-HCl, pH 8. One microgram of trypsin (Promega) in 10 mM IAA, 50 mM Tris-HCl, pH 8 was added and incubated for 1 h at 37 °C in dark. Then the peptides were eluted from SCX/SAX mixed beads with 60 μL 500 mM NaCl, pH 10 for 5 min. The eluted peptides were subsequently captured by C18 membrane and washed with 60 μL 0.1% (v/v) TFA for desalting. After the desalting step, the intact glycopeptides were eluted from C18 membrane to ZIC-HILIC beads with 60 μL 80% (v/v) ACN, 1% (v/v) TFA for 5 min. The tip was then washed with another 60 μL of 80% (v/v) ACN, 1% (v/v) TFA to remove non-specifically adsorbed peptides. Finally, intact glycopeptides were eluted with 60 μL 0.1% (v/v) TFA twice, lyophilized in vacuum, and store at -20℃ for LC-MS/MS analysis. Significantly, 1 μg trypsin is optimized for dealing with 1 μg protein sample. With the amount of protein sample increase, the amount of trypsin need increased to obtain better digestion efficiency, and 2 μg of trypsin were enough to process 25 μg protein sample.

Intact glycopeptide enrichment using traditional method

The proteins were dissolved in 8 M urea buffer containing 50 mM ABC, pH 8. After reduction and alkylation with 10 mM DDT and 20 mM IAA, the solution was then diluted to 1.5 M urea with 50 mM ABC. Then trypsin at a ratio of 1/40 (enzyme/protein, w/w) was added and digested overnight at 37 °C. After digestion, the digested peptides (25 μg) were desalted using a C18 SPE cartridge and lyophilized. For 1 μg starting amount, the peptides were desalted using homemade C18 StageTip method[2]. Then the lyophilized peptides were reconstituted in loading buffer 80% (v/v) ACN, 1% (v/v) TFA for intact glycopeptides enrichment using homemade ZIC-HILIC tip. Briefly, after ZIC-HILIC tip was conditioned with loading buffer, the peptides were loaded into the tip, repeated once and then washed with 60 μL loading buffer for three times. Finally, intact glycopeptides were eluted with 60 μL 0.1% (v/v) TFA for twice and subsequently lyophilized and store at -20℃ for LC-MS/MS analysis.

LC-MS/MS Analysis

The intact glycopeptides from standard sample IgG and fetuin were analyzed with Q Exactive MS equipped with easyLC system (Thermo Scientific). The sample was separated by a homemade C18 column with integrated tip (15 cm × 100 μm i.d.) at a flow rate of 250 nL/min with a 60 min gradient (5% B to 10% B for 3 min, increase to 40% B in 42 min, then increase to 55% B in 5 min, increase to 90% B in 1 min and hold for the last 9 min). Solvent A was H_2O containing 0.1% (v/v) FA and solvent B was ACN containing 0.1% (v/v) FA. The MS parameters were set according to previous reports with minor modification[3]. The MS spectra were acquired in the scan range (m/z) of 400 - 1800 with mass resolution of 70,000. For MS/MS acquisition, the first mass was fixed at 110.0 and the resolution was set as 17,500. Stepped normalized collision energy was set at 20%, 30%, 30.1%. The automatic gain control (AGC) was set as 500,000 and maximum injection time was set as 250 ms. Isolation window was set as 3.0 m/z. Dynamic exclusion was set as 20 s. The intact glycopeptides from KP4 cells were analyzed with Q Exactive HF-X MS (Thermo Scientific). The MS parameters were the same as above except for resolution was set 120,000 for MS1 and 15, 000 for MS/MS.

Data Analysis

Raw data was processed by pGlyco 2.0 for intact glycopeptide identification[4]. Mass tolerances were set as ± 10 ppm for precursors and ± 20 ppm for fragment ions. The enzyme was full-trypsin digestion with two missed cleavages. Cysteine carbamidomethylation (+57.021 Da) and methionine oxidation (+15.995 Da) were set as fixed modification and variable modifications, respectively. For analyzing intact glycopeptides from KP4 cells secretome, acetylation on protein N-term (+42.011 Da) was also set as variable modifications and the protein databases were download from SwissProt-human (released in January 2019; Homo sapiens, 20413 entries). The protein database contains all four IgG subclasses was used for analyzing IgG data: Uniprot ID of P01857 for IgG1, P01859 for IgG2, P01860 for IgG3 and P01861 for IgG4. The identified intact glycopeptides were filtered to ≤ 1% total FDR. DAVID (version 6.8) database was used to enriched Gene Ontology Molecular Function (GOMF) of the identified glycoproteins. The p value was set to ≤ 0.05.

Results and discussion

We went on to analyze the top 10 enriched terms of molecular function by GO analysis (Fig. S8). Proteins binding to transmembrane receptor integrin and extracellular collagen represent the top hits, which is consistent to the dense extracellular matrix component of pancreatic cancer tumor and their involvement in the modulation of cell adhesion and cell motility. In addition, the top hit proteins also participate to the signal transduction through binding to ligand or receptors, such as TGFβ and semaphorin receptor. Some of them are involved in the regulation of immune response and drug resistance of cancer cells through interaction with chemorepellent activity. These results are consistent to the general molecular features of pancreatic tumor microenvironment and should provide useful resource for further exploring their functional importance, especially from the aspect of glycosylation.

Supporting Figures

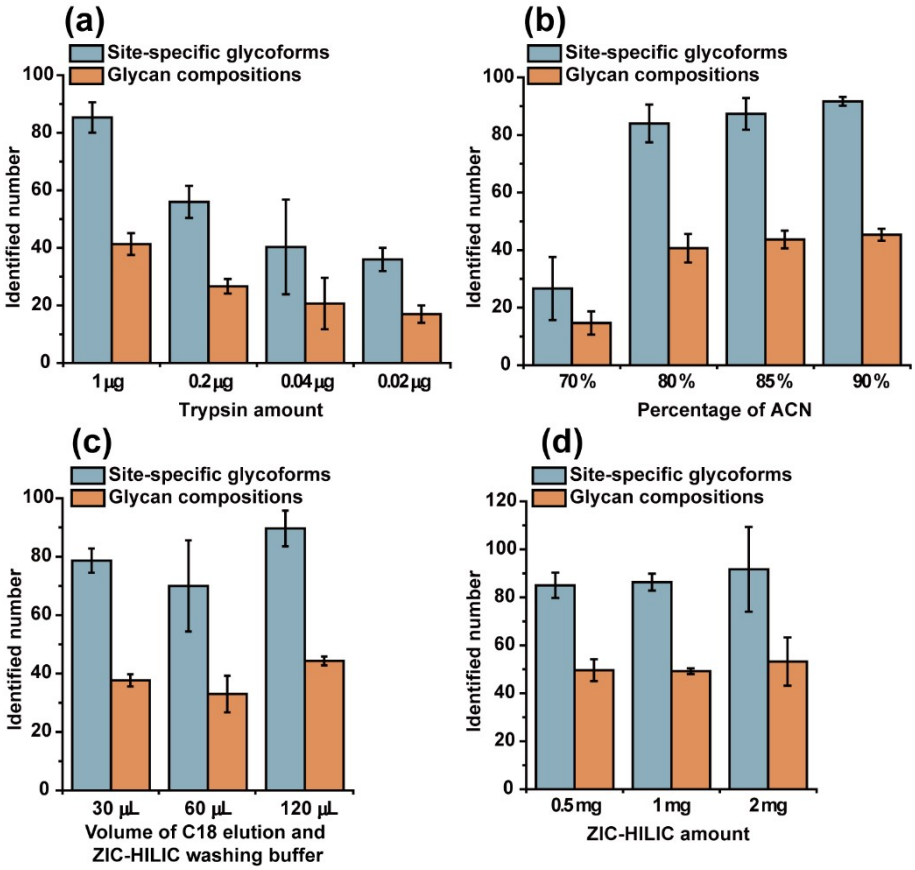

Fig. S1 Optimization of the Intact GlycoSISPROT for intact glycopeptide analysis starting from 1 μg IgG, including (a) amount of trypsin used in the digestion step, (b) concentration of ACN and (c) total volume of 80% (v/v) ACN,1% (v/v) TFA used in the peptide eluted and intact glycopeptide enrichment steps and (d) amount of ZIC-HILIC material used for intact glycopeptide enrichment.

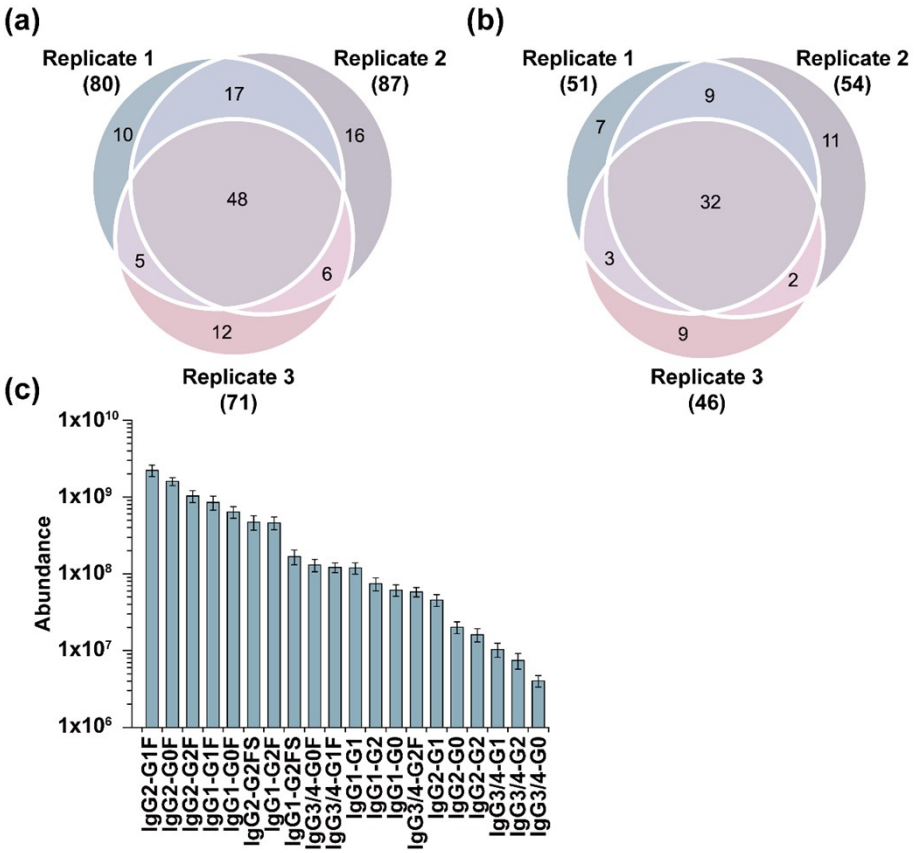

Fig. S2 Glycosylation analysis of 1 μg IgG by Intact GlycoSISPROT. Venn diagram of (a) site-specific glycoforms and (b) glycan compositions identified in three replicates. (c) Abundance (peak area of the MS1 elution profile) of 20 IgG intact glycopeptides were shown with error bar (n=3). IgG1-G0F indicates the peptide sequence EEQYNSTYR and the glycan composition Hex3HexNAc4Fuc1. IgG2 stands for peptide sequence EEQFNSTFR from IgG2; IgG3/4 stands for peptide sequence EEQYNSTFR or EEQFNSTYR from IgG3 or IgG4, respectively. G1F, G2F, G2FS, G0, G1 and G2 stand for Hex4HexNAc4Fuc1, Hex5HexNAc4Fuc1, Hex5HexNAc4Fuc1NeuAc1, Hex3HexNAc4, Hex4HexNAc4 and Hex5HexNAc4, respectively.

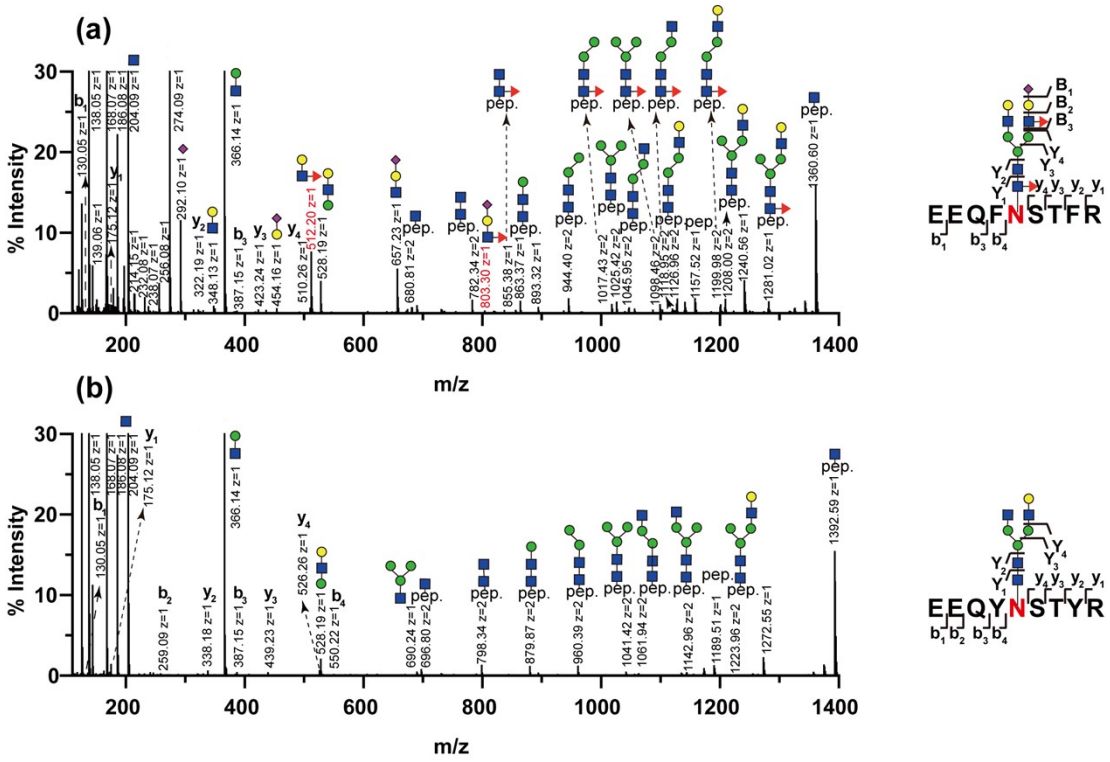

Fig. S3 MS/MS spectra of the intact glycopeptides from IgG (a) IgG2-G2F2S with precursor $[M + 3H]^{3+}$ at m/z 1121.78 and (b) IgG1-G1 with precursor $[M + 3H]^{3+}$ at m/z 884.02.

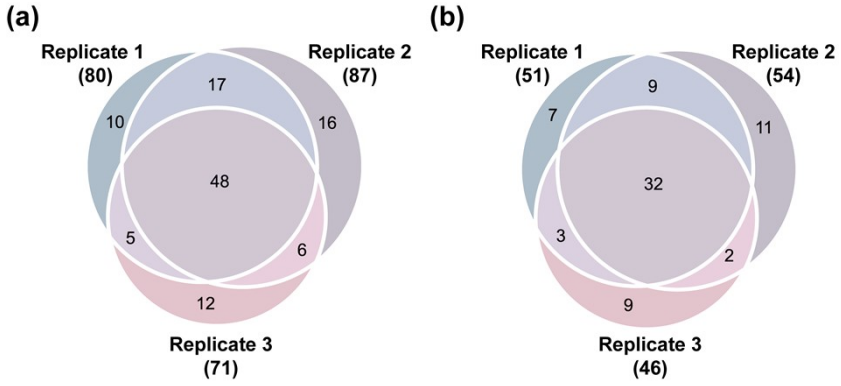

Fig. S4 Glycosylation analysis of fetuin by Intact GlycoSISPROT. Venn diagram of (a) site-specific glycoforms and (b) glycans identified in three replicates.

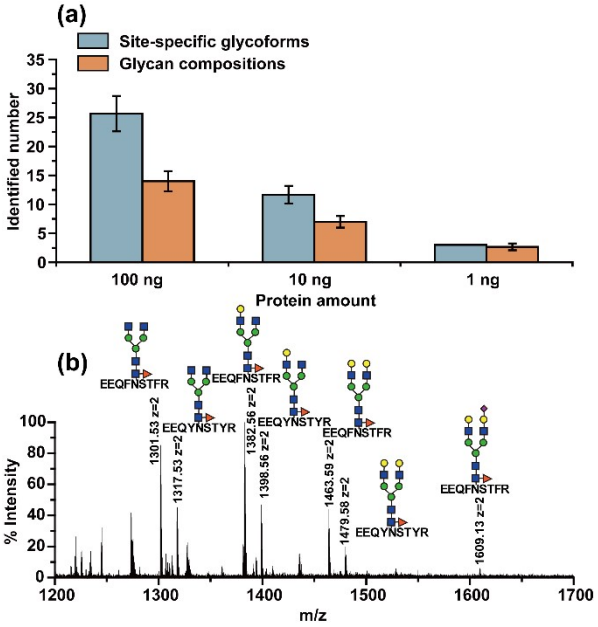

Fig. S5 Sensitivity of the Intact GlycoSISPROT. (a) Numbers of site-specific glycoforms and glycan compositions identified with different amount of IgG. (b) An average LC-MS spectrum of 1 ng IgG. The retention time was range from 10-25 min since most IgG intact glycopeptides eluted at this region. IgG1-G0F indicates the peptide sequence EEQYNSTYR and the glycan composition $Hex_3HexNAc_4Fuc_1$. IgG2 stands for peptide sequence EEQFNSTFR from IgG2. G1F, G2F and G2FS stand for $Hex_4HexNAc_4Fuc_1$, $Hex_5HexNAc_4Fuc_1$ and $Hex_5HexNAc_4Fuc_1NeuAc_1$, respectively.

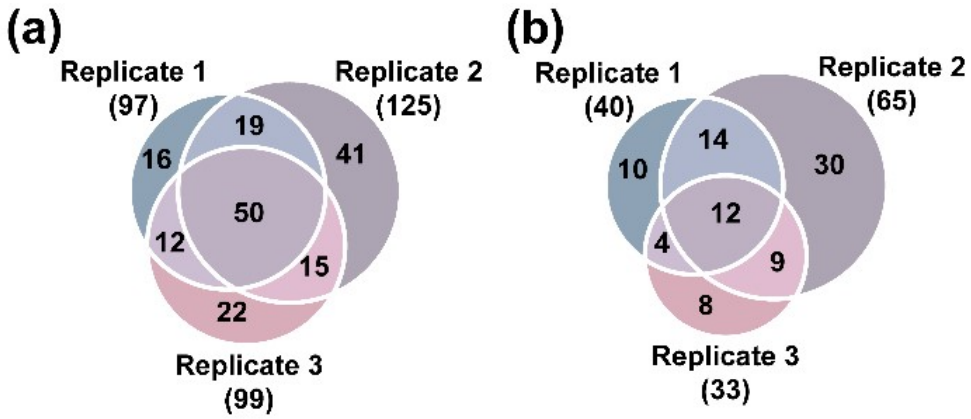

Fig. S6 Venn diagram of site-specific glycoforms identified from 1 μg of secreted proteins by (a) the Intact GlycoSISPROT and (b) traditional method in triplicate analysis.

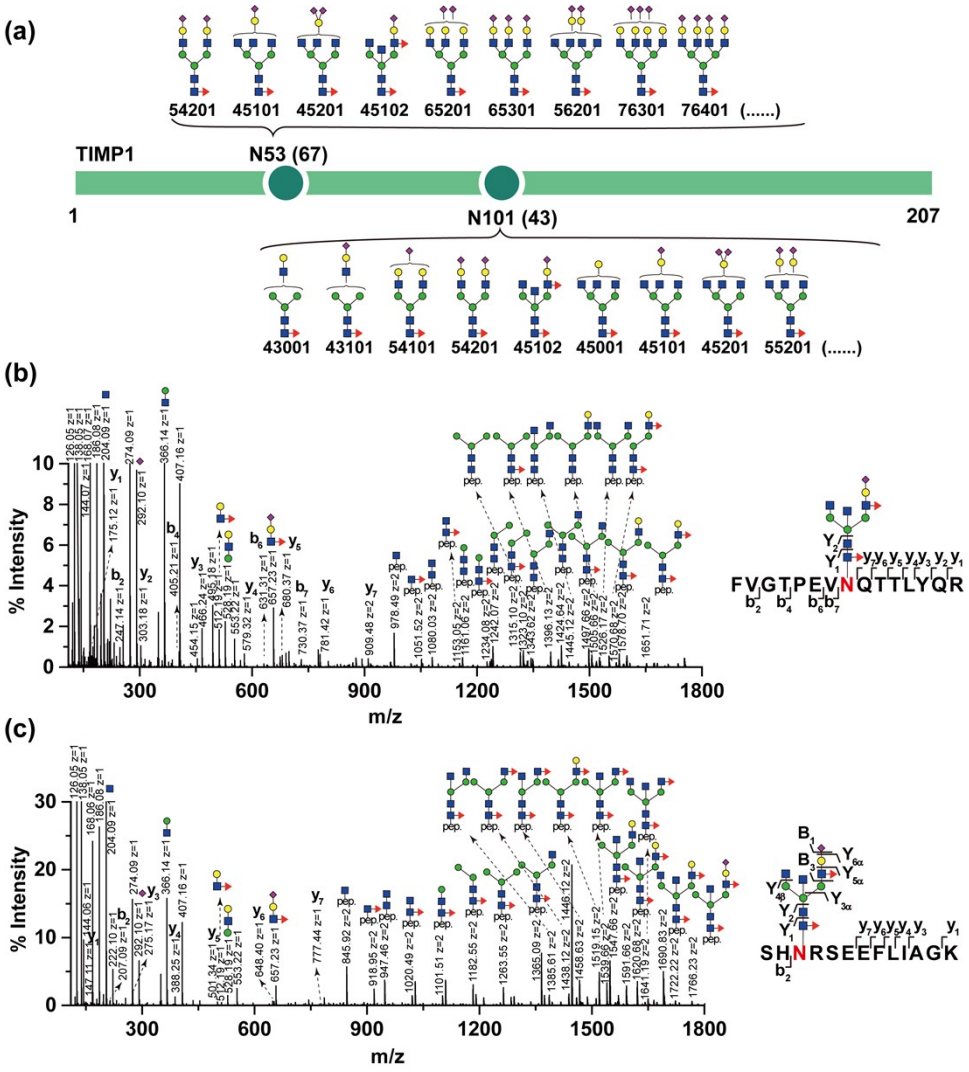

Fig. S7 Glycosylation heterogeneity of TIMP1. (a) Nine representative glycan compositions with the highest number of GPSMs identified at the two glycosites of TIMP1. (b-c) Annotated MS/MS spectra of identified intact glycopeptides with outer arm fucosylation from (b) N53, precursor $[M + 3H]^{3+}$ at m/z 1333.91 and (c) N101, precursor $[M + 3H]^{3+}$ at m/z 1245.53.

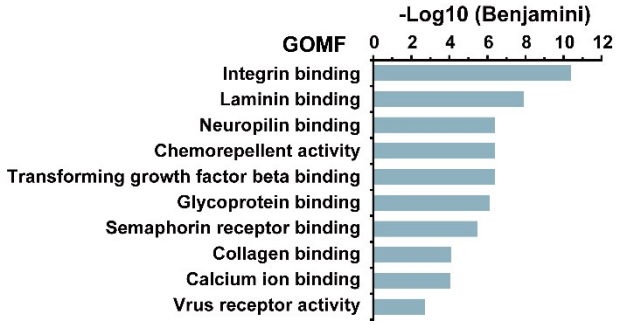

Fig. S8 Gene Ontology annotation for molecular function of glycoproteins identified from 25 µg of secreted proteins by the Intact GlycoSISPROT.

Reference

1 Y. Xue, J. J. Xie, P. Fang, J. Yao, G. Q. Yan, H. L. Shen, P. Y. Yang, *Analyst*, 2018, **143**, 1870-1880.

2 W. D. Chen, L. Chen, R. J. Tian, *Analyst*, 2018, **143**, 3693-3701.

3 Y. Wang, Z. Tian, *J. Am. Soc. Mass. Spectrom.*, 2020, **31**, 651-657.

4 M. Q. Liu, W. F. Zeng, P. Fang, W. Q. Cao, C. Liu, G. Q. Yan, Y. Zhang, C. Peng, J. Q. Wu, X. J. Zhang, H. J. Tu, H. Chi, R. X. Sun, Y. Cao, M. Q. Dong, B. Y. Jiang, J. M. Huang, H. L. Shen, C. C. L. Wong, S. M. He, P. Y. Yang, *Nat. Commun.*, 2017, **8**, 438.