A Chemoenzymatic Approach for the Proteomics Analysis of Mucin-type Core-1 O-glycosylation in Human Serum

Xin You[1, 2], Yating Yao[1, 2], Jiawei Mao[1, 2], Hongqiang Qin[1], Xinmiao Liang[1], Liming Wang[3], Mingliang Ye[*1]

[1] CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian, 116023, China

[2] University of Chinese Academy of Sciences, Beijing, 100049, China

[3] Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116023, China

Figure S1. MALDI spectrum of the standard O-GalNAcylated peptide before (A) ([M+H]+=1191.71 Da, [M+Na]+=1213.74 Da, [M+K]+=1229.74 Da) and after (B) ([M+Na]+=1211.47 Da, [M+K]+=1227.43 Da) oxidated by galactose oxidase.

Figure S2. (A) The number of unique O-GalNAcylated peptide sequences and proteins from Jurkat cell lines digest identified by HILIC enrichment and chemoenzymatic enrichment. (B) The overlap of identified O-GalNAc modified peptide sequences and proteins by chemoenzymatic method (blue circle) and HILIC method (yellow circle). (C) The cellular components analysis of identified O-glycoproteins by chemoenzymatic method (orange column) and HILIC method (green column) with p-Value<0.05.

Figure S3. The MS/MS spectrum of false positive identified O-glycopeptides in HILIC enrichment method. (A) N-glycopeptide was falsely identified as an O-glycopeptide. The presence of oxonium ions indicates it is a glycopeptide, the presence of the NX(S/T) motif indicates it is an N-glycopeptide. This could be due to the incomplete deglycosylation of N-glycosylation. (B) Non-glycodpeptide was falsely identified as an O-glycopeptide. The absence of oxonium ions indicates it is not a glycopeptide. The poor specificity of the HILIC enrichment leaded to the increase of random matches.

Figure S4. The MS/MS spectra of four O-glycopeptides identified with low Xcorr.

Figure S5. (A) The MS/MS spectrum of (VQAAVGTSAAPVPSDNH +GalNAc) containing the reported ions of m/z=171.0766 Da and 195.0763 Da. (B) The MS/MS spectrum of 171.07 of (A). The intensity of the peak 171.0766 is around 40000 while the intensity of noise in 964.97. Therefore the calculated S/N ratio of the peak 171.0766 was over 40. This high S/N ratio indicated that this peak 171.0766 was an

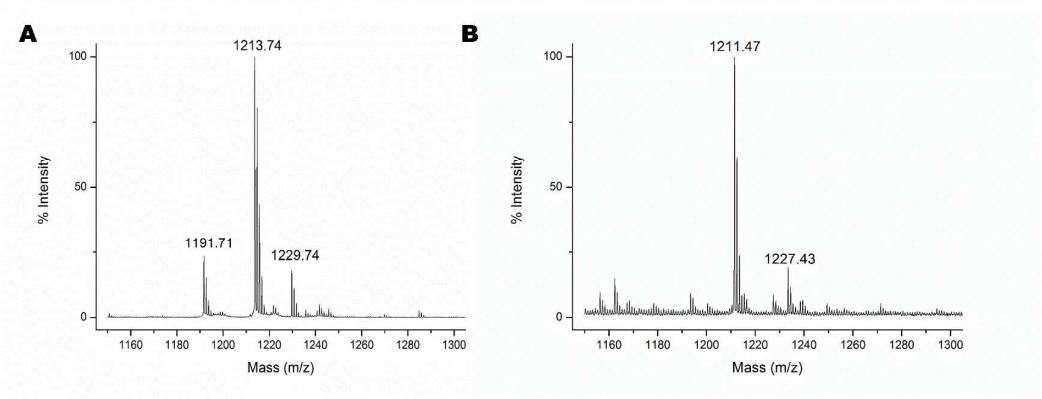
individual peak instead of noise. ("N" means the intensity of noise calculated by the Xcalibre).

Figure S6. The (A) MS/MS spectrum of O-glycopeptide (TTPPTTATPIR+GalNAc+Galβ1-3GalNAc) and (B) the low mass range of this MS/MS spectrum.

Figure S7. The O-GalNAcyalted range (marked with red column) on IGHG3_HUMAN identified by our method.

Figure S8. The overlap between this work (blue circle) and our previous work[3] (yellow circle) of 1D analysis at O-GalNAc modified (A) peptide sequences level and (B) proteins level. (The number of identified unique O-GalNAc modified peptide sequences, proteins and the ratio was marked in the Venn figure)

The following tables are given as separate Excel files:
Table S1. Identified O-glycosylation in Jurkat cell by HILIC and GAO
Table S2. Identified O-glycosylation in bovine fetuin-A native and deSA
Table S3. Identified O-glycosylation in human serum NCE25
Table S4. Identified O-glycosylation in human serum NCE28
Table S5. Identified O-glycosylation in human serum NCE31
Table S6. Identified O-glycosylation in human serum NCE34



Peptide Sequence: T*VPAAVVVA-OH
Glycan: β-Gal (1-3) α-GalNAc

Figure S1. MALDI spectrum of the standard O-GalNAcylated peptide before (A) ($[M+H]^+$=1191.71 Da, $[M+Na]^+$=1213.74 Da, $[M+K]^+$=1229.74 Da) and after (B) ($[M+Na]^+$=1211.47 Da, $[M+K]^+$=1227.43 Da) oxidated by galactose oxidase.

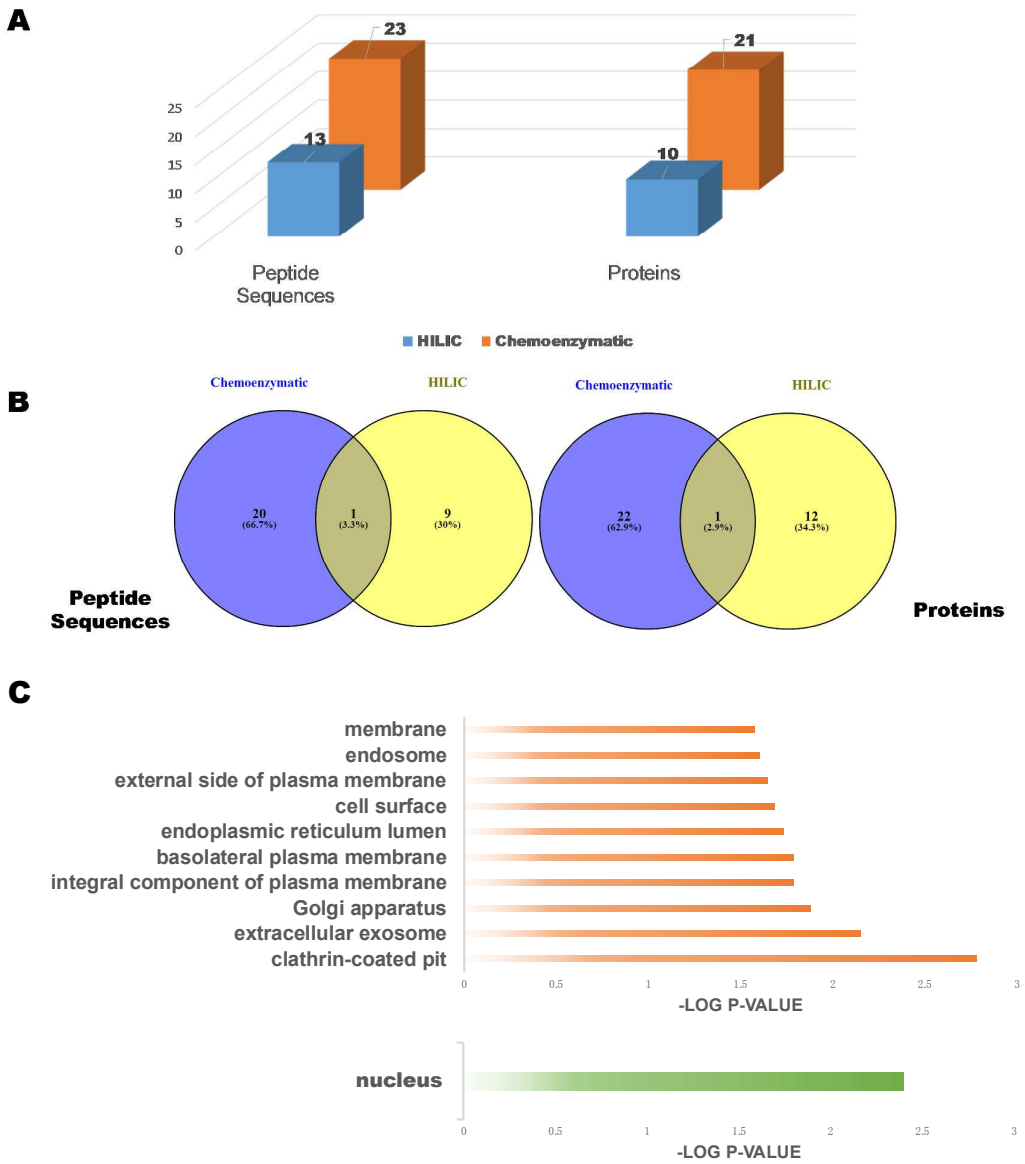


Figure S2. (A) The number of unique O-GalNAcylated peptide sequences and proteins from Jurkat cell lines digest identified by HILIC enrichment and chemoenzymatic enrichment. (B) The overlap of identified O-GalNAc modified peptide sequences and proteins by chemoenzymatic method (blue circle) and HILIC method (yellow circle). (C) The cellular components analysis of identified O-glycoproteins by chemoenzymatic method (orange column) and HILIC method (green column) with p-Value<0.05.[1,2]

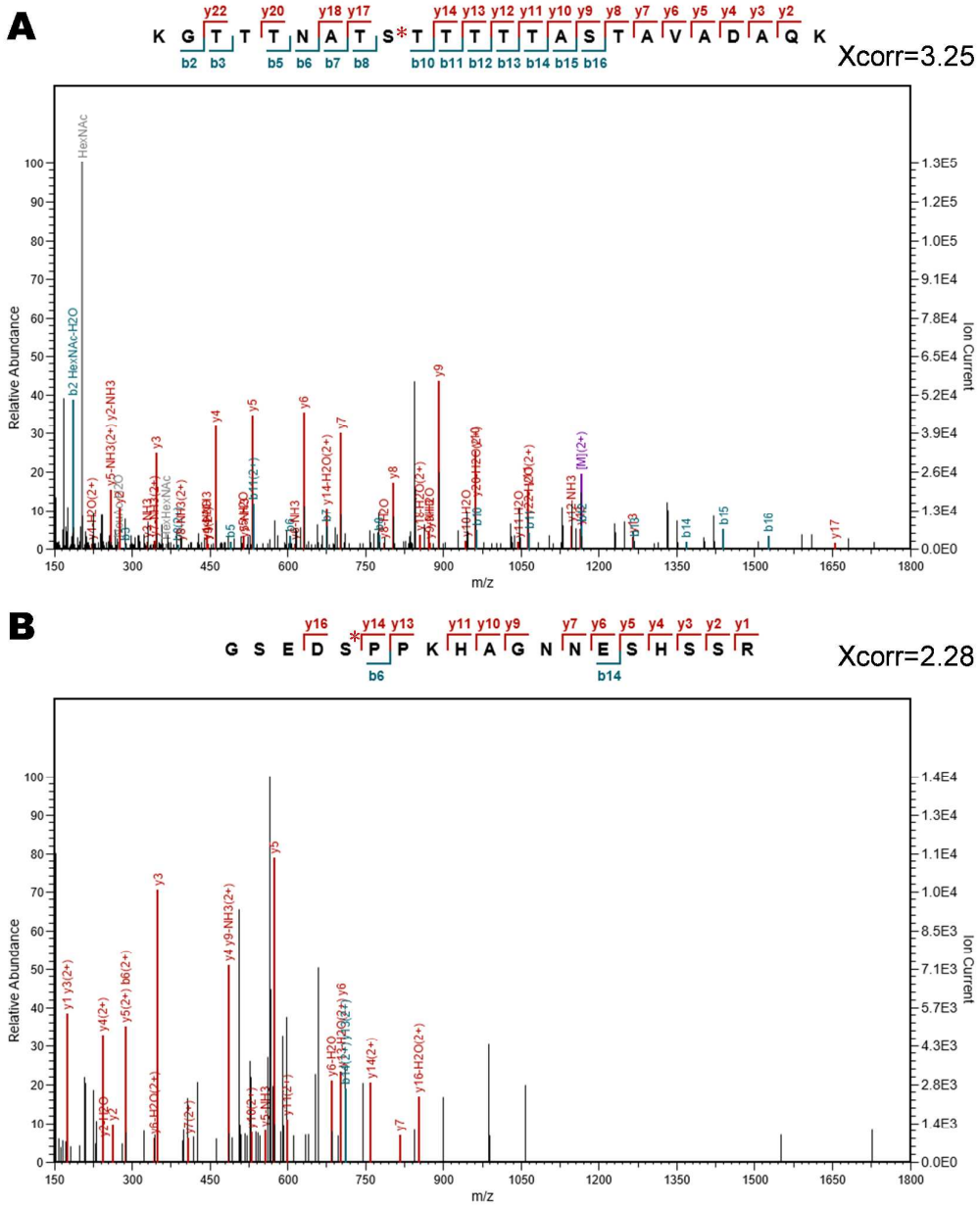


Figure S3. The MS/MS spectrum of false positive identified O-glycopeptides in HILIC enrichment method. (A) N-glycopeptide was falsely identified as an O-glycopeptide. The presence of oxonium ions indicates it is a glycopeptide, the presence of the NX(S/T) motif indicates it is a N-glycopeptide. This could be due to the incomplete deglycosylation of N-glycosylation. (B) Non-glycodpeptide was falsely identified as an O-glycopeptide. The absence of oxonium ions indicates it is not a glycopeptide. The poor specificity of the HILIC enrichment leaded to the increase of random matches.

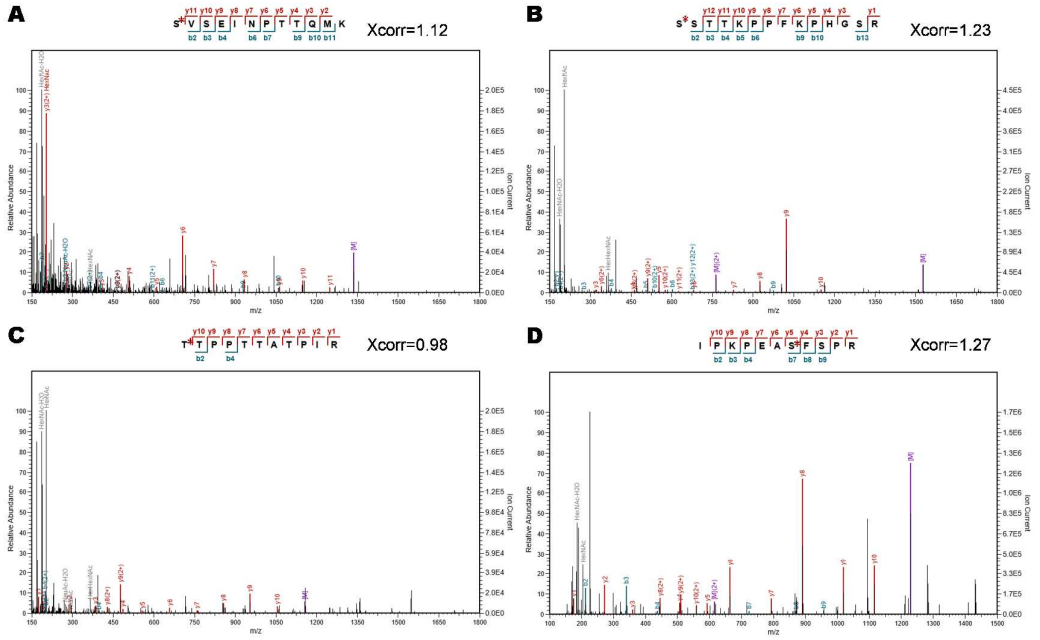


Figure S4. The MS/MS spectra of four O-glycopeptides identified with low Xcorr.

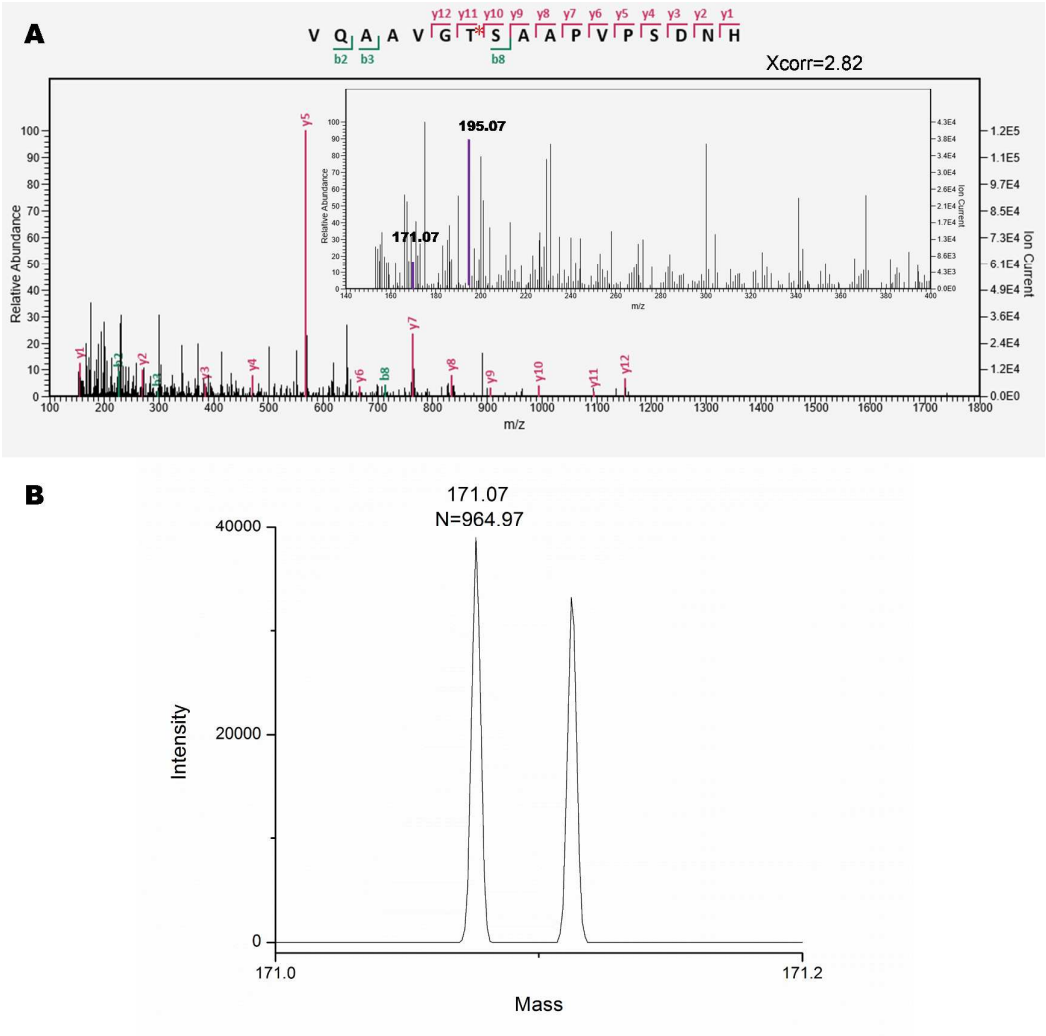


Figure S5. (A) The MS/MS spectrum of (VQAAVGTSAAPVPSDNH +GalNAc) containing the reported ions of m/z=171.0766 Da and 195.0763 Da. (B) The MS/MS spectrum of 171.07 of (A). The intensity of the peak 171.0766 is around 40000 while the intensity of noise in 964.97. Therefore the calculated S/N ratio of the peak 171.0766 was over 40. This high S/N ratio indicated that this peak 171.0766 was an individual peak instead of noise. ("N" means the intensity of noise calculated by the Xcalibre).

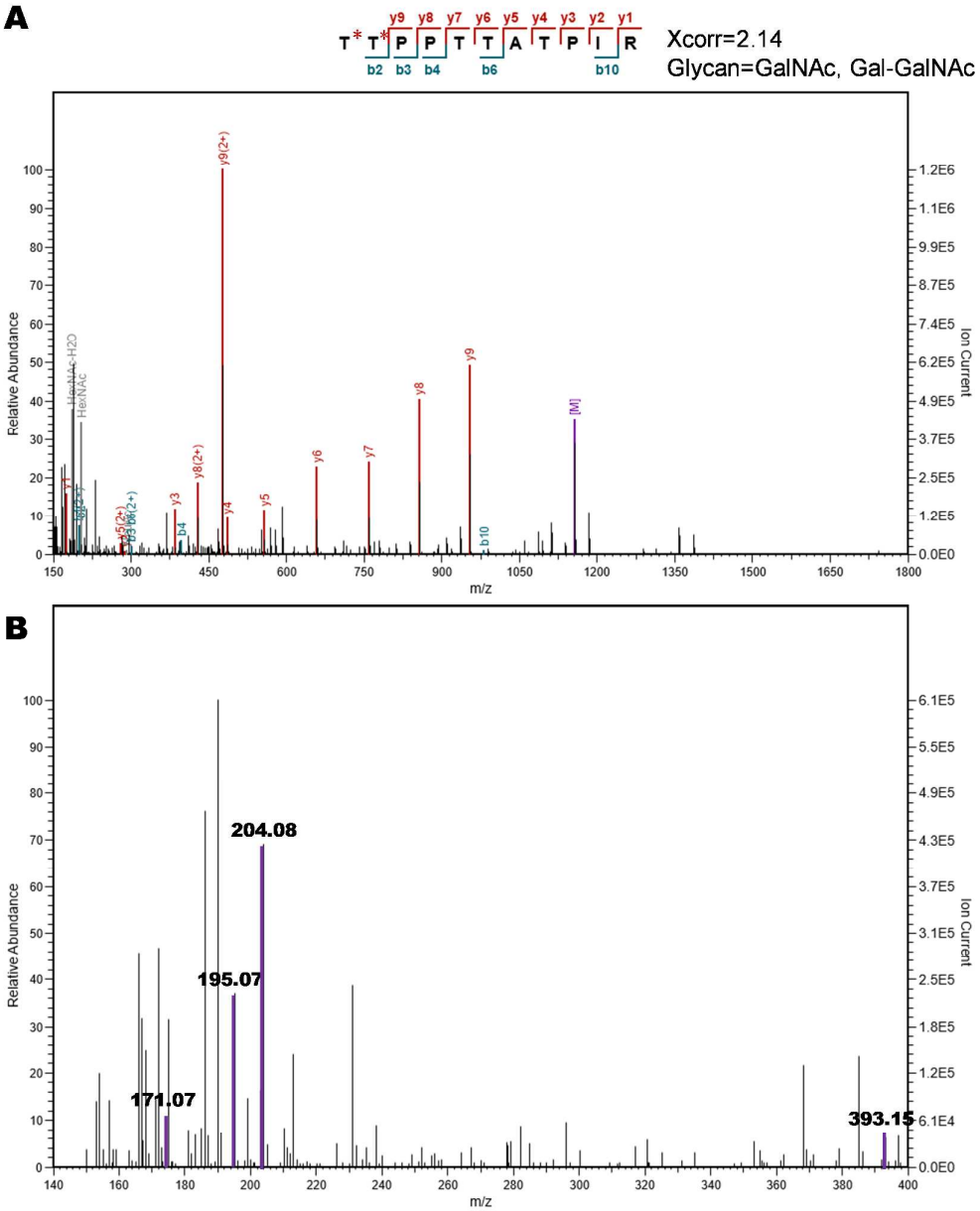


Figure S6. The (A) MS/MS spectrum of O-glycopeptide (TTPPTTATPIR+GalNAc+Galβ1-3GalNAc) and (B) the low mass range of this MS/MS spectrum.

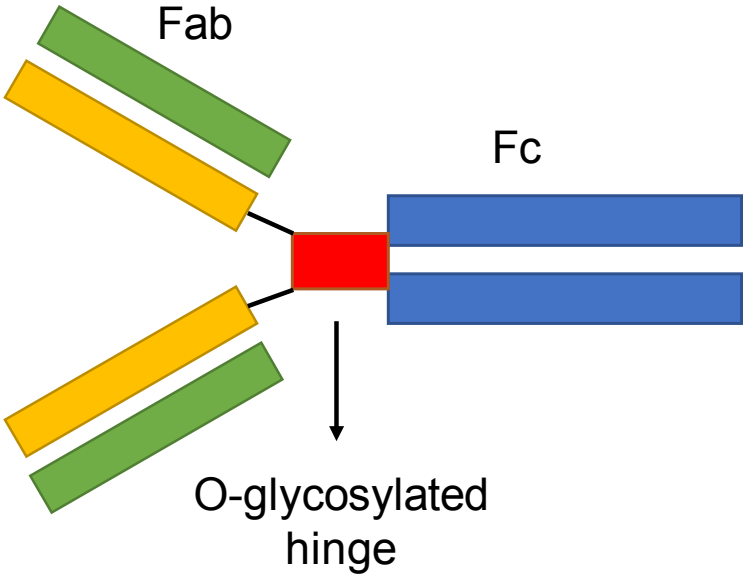


Figure S7. The O-GalNAcyalted range (marked with red column) on IGHG3_HUMAN identified by our method.

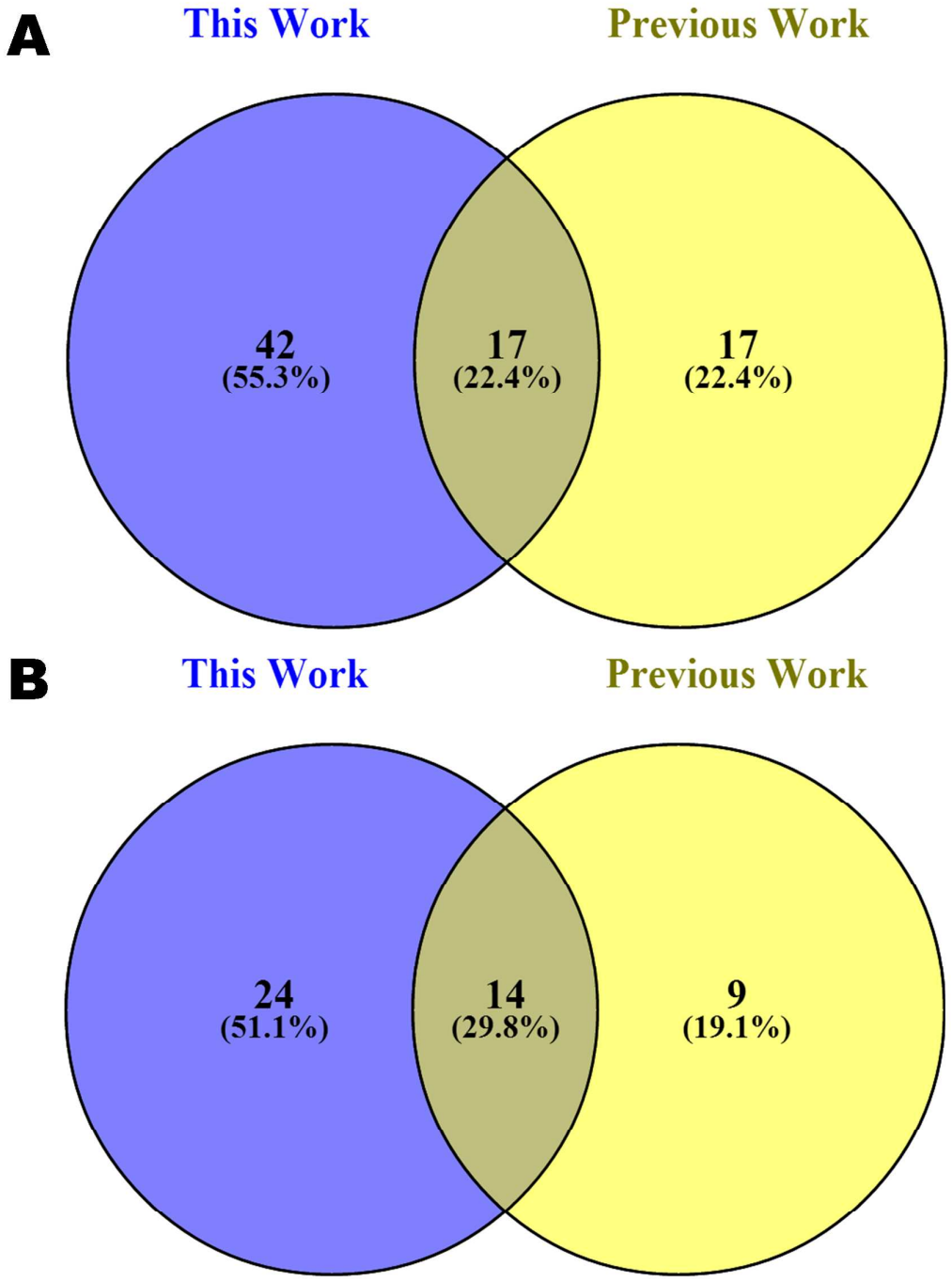


Figure S8. The overlap between this work (blue circle) and our previous work[3] (yellow circle) of 1D analysis at O-GalNAc modified (A) peptide sequences level and (B) proteins level. (The number of identified unique O-GalNAc modified peptide sequences, proteins and the ratio was marked in the Venn figure)

References:

(1) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nature protocols* **2009**, *4*, 44-57.

(2) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic acids research* **2009**, *37*, 1-13.

(3) Qin, H.; Cheng, K.; Zhu, J.; Mao, J.; Wang, F.; Dong, M.; Chen, R.; Guo, Z.; Liang, X.; Ye, M.; Zou, H. *Proteomics Analysis of O-GalNAc Glycosylation in Human Serum by an Integrated Strategy, Analytical chemistry* **2017**, *89*, 1469-1476.