



HHS Public Access

Author manuscript

Nat Protoc. Author manuscript; available in PMC 2021 November 26.

Published in final edited form as:

Nat Protoc. 2020 August ; 15(8): 2589–2610. doi:10.1038/s41596-020-0345-1.

Large-scale Mapping of Site-specific O-GalNAc Glycoproteome

Weiming Yang^{*}, Angellina Song, Minghui Ao, Yuanwei Xu, Hui Zhang

Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Abstract

Protein O-GalNAcylation is ubiquitous in many proteins that reside in the secretory pathway, cell surface, and extracellular spaces. As a major type of protein glycosylation, O-GalNAcylation attaches GalNAc-type glycans to protein Ser and Thr residues via an O-linked glycosidic bond. Thus, a massive number of Ser and Thr residues in the human proteome sequence could be potential sites of O-GalNAcylation. O-GalNAcylation supports folding, stability, trafficking, and functions of carrier proteins. Identifying site-specific O-GalNAc glycoproteome is prerequisite to discover the biological significance of this modification in diseases. However, lack of suitable methodology, absence of consensus sequon of O-GalNAcylation sites, and complex O-GalNAc glycan structures pose analytical hurdles. It is particularly the case for the analysis of clinical samples, including tissues, blood, body fluids, and primary cells, whose site-specific O-GalNAc glycoproteome remains mostly unknown to date. We recently reported the development of mass spectrometry (MS)-based method named EXoO that enabled large-scale mapping of O-GalNAcylation sites in the clinical samples. EXoO has a unique strength to map dense O-GalNAcylation sites clustered in vincinal locations in mucin-type glycoproteins that are not easily studied using other methods. The protocol is virtually applicable to study all glycoproteins since glycoproteins in the locations from the Golgi to the plasma membrane to extracellular space could be modified by O-GalNAcylation. Here, a detailed protocol of EXoO is described that includes eight steps: (i) extraction and proteolytic digestion of proteins to peptides, (ii) sequential guanidination and desalting of peptides using C18 cartridge, (iii) enrichment of glycopeptides using RAX cartridge, (iv) conjugation and release of site-specific O-GalNAc glycopeptides from solid phase, (v) preparation of O-GalNAc glycopeptides for LC-MS/MS analysis, (vi) LC-MS/MS analysis, (vii) database search, and (viii) data filtering and reporting. Thousands of O-GalNAcylation sites from hundreds of glycoproteins with information regarding site-specific O-GalNAc glycans can be determined in a complex sample. The protocol can be performed by a researcher with basic skills for proteomic experiment and takes ~ four days to complete.

^{*} **Corresponding Author: Address:** Department of Pathology, Johns Hopkins University School of Medicine, 400 North Broadway, Room 4001A, Baltimore, Maryland, United States. wyang21@jhmi.edu.

AUTHOR CONTRIBUTIONS

W.Y. and H.Z. conceived the research. W.Y. A. S., and Y.X. conducted the experiments and validated the protocol. W.Y., Y.X. and M.A. conducted data analysis.

Code availability

Extract_oxonium_ion_spectra_intensity.exe script is free available at GitHub (download at <https://github.com/appleTree2017/Extraction-of-O-linked-glycopeptides-EXoO->)

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

INTRODUCTION

Transmembrane proteins or proteins that are expressed on the extracellular surface are the interfaces between the cell interior and the outside environment^{1,2}. They are easily accessible to therapeutic drugs³, ligands⁴, immunotherapy⁵, and CAR-T cell therapy⁶ and are crucial to infection by pathogenic bacteria⁷ and viruses⁸. These features make cell surface proteins a highly interesting class of proteins for biology and clinical research. Also, proteins secreted in body fluids such as serum, cerebrospinal fluid, urine, etc. are readily accessible and provide a detailed window into the state of health of an individual⁹. These two classes of protein are, therefore, useful for patient diagnosis, clinical management, and therapeutics. A key feature of the proteins described above is that a large fraction of them are modified by O-GalNAcylation¹. Glycoproteins decorated with O-GalNAc glycans are complex with each potentially containing multiple O-GalNAcylation sites and each site containing different glycans¹⁰. The O-GalNAc glycans on these glycoproteins form a cell surface glycocalyx that plays crucial roles in eukaryotic development, cell signaling, cell-cell interactions, cell adhesion, receptor-ligand binding, drug/therapy accessibility, and interactions with pathogenic bacteria and viruses¹⁰. Important O-GalNAc glycoproteins include different immunoglobulins^{11,12}, erythropoietin (EPO)¹³, HIV Env¹⁴, mucin proteins¹⁵, and cluster of differentiation (CD) proteins¹⁶, etc.

Owing to lack of effective methodology, the absence of consensus motif, and complex structure of O-GalNAc glycans, large-scale mapping of O-GalNAcylation sites in the complex samples is highly challenging, particularly for the analysis of clinical samples including tissues, primary cells, blood, and other body fluids, whose site-specific O-GalNAc glycoproteome remains elusive to date. Among the current methods for mapping of O-GalNAcylation sites, five main methods have been reported: SimpleCell¹⁷, Isotag¹⁸, hydrazide capture¹⁹, lectin enrichment²⁰, and hydrophilic interaction chromatography (HILIC) enrichment²¹. These methods need mass spectrometry (MS) using electron transfer dissociation (ETD) for the localization of O-GalNAcylation sites and the definition of their site-specific O-GalNAc glycans in proteins, with only a limited number of sites being mapped in clinical samples. As a result, the landscape of O-GalNAc glycoproteome has been significantly underestimated, and consequently, the significance of O-GalNAc glycoproteome in diseases has not been well-appreciated. This all indicates that a broadly applicable method, independent of ETD-MS, for large-scale identification of site-specific O-GalNAc glycoproteome and defining the corresponding glycans particularly if it can be applied for clinical samples would represent a groundbreaking advance towards achieving an understanding of O-GalNAcylation.

In the present protocol, we describe an MS-based method for the site-specific extraction of O-linked glycopeptides (EXoO), which enables simultaneous enrichment and identification of both the O-GalNAcylation sites and site-specific glycans in unprecedented depth¹⁶. EXoO is highly effective for studying clinical samples¹⁶. In sharp contrast to other existing methods that report the mapping of at most a few hundred glycosylation sites, the use of EXoO method maps over 3,000 O-GalNAcylation sites¹⁶. The superior performance of EXoO benchmarks it as a highly effective method for revealing the landscape of the site-specific O-GalNAc glycoproteome¹⁶. Also, EXoO has a unique strength to study mucin-type

glycoproteins¹⁶. Mucin-type glycoproteins contain dense O-GalNAc glycans clustered in vincinal sites that hinder digestion by conventional proteases for analysis¹⁵. OpeRATOR enzyme in the EXoO method is naturally created to cleave and enable localization of O-GalNAcylation sites in mucin-type glycoproteins¹⁶. Therefore, EXoO is particularly useful to study mucin-type glycoproteins in certain diseases such as cancer²² and eukaryotic development²³. EXoO is used by an independent laboratory to study a few recombinant glycoproteins²⁴.

Development of the protocol

EXoO tackles the unmet need of large-scale mapping of O-GalNAcylation sites in the clinical samples that have hardly been achieved in the last decades. As a result, the important roles of O-GalNAc glycoproteins in different biological systems remain concealed. Our group focuses on solid phase approach to study N-linked glycoproteome^{25,26}, N-, and O-linked glycans²⁷. The use of solid phase allows extensive washes to remove unwanted peptides and impurities. Glycopeptides or glycans can be selectively released from the solid phase using enzymes or chemicals^{16,25,26,28}. In the case of EXoO, OpeRATOR is used to releasing site-containing O-GalNAc glycopeptides from the solid phase¹⁶. The OpeRATOR is an endoprotease discovered from mucin-digesting human gut bacterial *Akkermansia muciniphila*¹⁶. The uniqueness of OpeRATOR is that it recognizes O-GalNAc glycan followed by cleavage at N-termini of glycan-occupied Ser and Thr¹⁶. The released O-GalNAc glycopeptides have the glycosylation sites at the N-termini of peptide sequences, and the glycans are still attached to the peptide backbones¹⁶. The enzymatic feature of OpeRATOR is well-suited with our solid phase approach to allow capture and specific release of site-containing O-GalNAc glycopeptides with high purity for LC-MS/MS analysis¹⁶. The protocol was used to profile site-specific O-GalNAc glycoproteome in human kidney tissue, serum, and T cells¹⁶. Over 3,000 O-GalNAcylation sites with the definition of their site-specific glycans on over 1,000 proteins are mapped that include 1,781, 732 and 1,295 O-GalNAcylation sites in kidney tissue, serum and T cells, respectively¹⁶. Furthermore, the use of EXoO reveals significant differences of O-GalNAc glycoproteome between tumor and normal kidney tissues demonstrating its applicability in the clinical study¹⁶. Of the changed glycoproteins, EXoO shows its unique strength to map O-GalNAcylation sites and define site-specific glycans in mucin-type glycoproteins¹⁶. Versican core protein (VCAN) and aggrecan core protein (ACAN), also known as proteoglycans, show significant elevation spanning their 163 and 82 O-GalNAcylation sites, respectively¹⁶. Given the effectiveness of EXoO method in mapping site-specific O-GalNAc glycoproteome in the complex samples, the importance of O-GalNAcylation in different biological systems can be studied with site-specific precision.

Overview of the technique

The protocol is summarized in Figure 1a&b and includes the following eight steps:

- In the first step, proteins are extracted, denatured, reduced, alkylated, and digested using trypsin to generate peptides.
- In the second step, peptides are sequentially guanidinated and desalted on the same C18 cartridge.

Guanidination modifies primary amine group on the side chain of lysine residues to homoarginine to prevent conjugation of lysine residues to aldehyde groups on solid phase and enable identification of glycopeptides containing lysine residues²⁵.

- Enrichment of glycopeptides using RAX cartridge²⁹.

The enrichment of glycopeptides facilitates OpeRATOR digestion in a higher concentration and a smaller reaction volume in the downstream step.

- Conjugation of peptides to solid phase and release of site-containing O-GalNAc glycopeptides.

The use of solid phase allows maximal removal of impurities and contaminated peptides. Subsequent use of OpeRATOR releases site-containing O-GalNAc glycopeptides with high purity.

- Preparation of site-containing O-GalNAc glycopeptides for LC-MS/MS.

Glycopeptides are collected, desalted, and resuspended for analysis.

- LC-MS/MS analysis.

Site-containing O-GalNAc glycopeptides are analyzed using LC-MS/MS. In the protocol, HCD-MS2 is used since it is found to be efficient for peptide identification¹⁶.

- Database search.

Commercial software SEQUEST in Proteome Discoverer (v2.2) is used in the protocol. Alternatively, other proteomics or glycoproteomics search engines may be useful for the search.

- Data filtering and reporting.

Filtering criteria are designed after the study of MS/MS spectra of site-containing O-GalNAc glycopeptides¹⁶.

Advantages and limitations

EXoO is a highly effective method for large-scale mapping of O-GalNAcylation sites in the complex samples. Advantages of EXoO include (1) large-scale mapping of O-GalNAcylation sites and defining of site-specific glycans; (2) applicable to study clinical samples; (3) no need for ETD-MS2 for site localization; (4) suitable for the analysis of mucin-type O-GalNAc glycoproteins. To map O-GalNAcylation sites, other methods would require the use of ETD, an MS-based peptide fragmentation technique³⁰. ETD is efficient to fragment glycopeptides with high charge states and certain mass for site localization^{31–33}. EXoO does not require ETD but takes advantage of OpeRATOR that generate O-GalNAc glycopeptides with the glycosylation sites presenting at the N-termini of peptide sequences. Mucin-type glycoproteins contain dense clusters of O-GalNAcylation sites in vincinal positions that are covered by glycans to resist digestion by conventional proteases^{34,35}. OpeRATOR enzyme can cleave the O-GalNAc glycan-occupied sites to allow

O-GalNAcylation site localization in mucin-type glycoproteins¹⁶. EXoO thus is useful to study mucin-type glycoproteins in diseases.

Definition of site-specific O-GalNAc glycans may be cofounded in the case that the identified O-GalNAc glycan composition may consist of more than one glycan structure. For instance, an identified glycan composition Hex(2)HexNAc(2) may consist of a single core 2 structure or two Hex(1)HexNAc(1) core 1 structure. In our data, 69% of assigned PSM contains a single glycan composition Hex(1)HexNAc(1), most likely Gal(1)GalNAc(1) glycan structure at the O-GalNAcylation sites (Fig. 1c)¹⁶. The glycan specificity of OpeRATOR is apparent for Gal(1)GalNAc(1) but not for other types of O-GalNAc glycans¹⁶. Therefore, the protocol may not be applicable to identify O-GalNAcylation sites with glycans other than Gal(1)GalNAc(1). Definite establishment of glycan specificity of OpeRATOR merits future investigation. As recommended by the company, OpeRATOR is concurrently used with SialEXO, which removes sialic acid to facilitate digestion by OpeRATOR¹⁶. Hence, EXoO may not be useful to define site-specific O-GalNAc glycan containing sialic acid.

Applications of the method

In our application, human kidney tissue, serum, and T cells are analyzed that demonstrate the applicability of EXoO to analyze different types of clinical specimens¹⁶. O-GalNAcylation is ubiquitous in human that EXoO serves as a highly effective tool to characterize the O-GalNAc glycoproteome in a site-specific manner. Clinical samples with different diseases can be analyzed using EXoO to reveal the role of O-GalNAc glycoproteome. Since O-GalNAc glycoproteins are commonly present on the secretory pathway from ER to extracellular space, EXoO is useful to identify important O-GalNAc glycoprotein on the cell surface and in the secretion. EXoO is useful to characterize site-specific O-GalNAc glycoproteome in the whole human proteome. Establishment of O-GalNAc glycoproteome in the human proteome will provide a vital foundation to reveal important O-GalNAc glycoproteins in diseases. O-GalNAcylation also presents in animals that can be studied by EXoO. A unique feature of EXoO is its applicability to analyze mucin-type glycoproteins. An important mucin-type glycoprotein is IgA that has a mucin domain with several O-GalNAcylation sites being located closely. The mucin domain of IgA is not digested by common proteases that hampers acquisition of data. The OpeRATOR can digest these mucin-type domains to gain information regarding O-GalNAcylation sites and site-specific glycans. EXoO would be useful to characterize pharmaceutical glycoproteins with mucin domain.

Expertise needed to implement the protocol

The protocol involves the operation of MS instrument and LC-MS/MS data analysis. An MS core facility will be able to advise and assist the setup and analysis using MS. User may refer to software User Guide for using SEQUEST in Proteome Discoverer software.

MATERIALS:**REAGENTS**

- Urea (Sigma, cat. no. U0631–1KG)
- BCA protein assay reagent (Thermo Fisher Scientific, cat. no. 23225)
- 5M NaCl (Santa Cruz Biotechnology, cat. no. sc-295833)
- 1M Trizma® hydrochloride solution pH 8.0 (Tris-HCl pH 8.0; Sigma, cat. no. T2694–1L)
- 1M Trizma® hydrochloride solution pH 7.4 (Tris-HCl pH 7.4; Sigma, cat. no. T2663–1L)
- PBS pH 7.4 (1×; Thermo Fisher Scientific, cat. no. 10010049)
- PBS pH 7.4 (10×; Quality Biological, cat. no. 119-069-131)
- DTT (Thermo Fisher Pierce, cat. no. 20291)
- Iodoacetamide (IAA; Sigma, cat. no. I1149–5G)
- Ammonium hydroxide (NH₄OH; 28–30% (vol/vol); Sigma, cat. no. 338818–100ML) ! CAUTION NH₄OH is hazardous and flammable and should be handled in a hood using gloves.
- O-Methylisourea hemisulfate salt (Aldrich, cat. no. 455598–100G)
- Triethylammonium acetate buffer (1 M pH 7.0; Sigma, cat. No. 90358–500ML)
- Acetonitrile (ACN, Thermo Scientific, cat. no. A9554) ! CAUTION ACN is corrosive and should be handled in a hood using gloves
- Trifluoroacetic acid (TFA; Sigma, cat. no. T6508–1L) ! CAUTION TFA is a corrosive acid, and it should be handled in a hood using protective goggles and gloves.
- Deionized (DI) water (laboratory-made, or Thermo Fisher Scientific, cat. no. 1523001)
- Trypsin (Promega, cat. no. V5111)
- BSA protein standard (2 mg/ml; Thermo Fisher Scientific, cat. no. 23209)
- BCA protein assay reagent (Thermo Fisher Scientific, cat. no. 23225)
- Sodium cyanoborohydride (NaCNBH₃; Sigma-Aldrich, cat. no. 156159)
- Sodium chloride solution (NaCl; 5 M; ChemCruz Biochemicals, cat. no. SC-203274)
- AminoLink™ Plus coupling resin (Thermo Fisher Scientific, cat. no. 20501)
- OpeRATOR (Genovis, cat. no. G2-OP1–020)
- SialEXO (Genovis, cat. no. G2-OP1–020)

EQUIPMENT

- Eppendorf Research Plus pipette (Sigma-Aldrich, cat. no. Z683884-1EA)
- BD Falcon tube (15 ml; BD Biosciences, cat. no. 352095)
- Snap-cap spin column (1 ml; Thermo Fisher Scientific, cat. no. 69725)
- Corning® 96-well Clear Flat Bottom Polystyrene Not Treated Microplate, 20 per Bag, with Lid, Sterile (96-well plate; Corning, cat. no. 3370)
- Waters tC18 SepPak (Waters, cat. no. WAT054925)
- HYPERSEP Retain-AX 500MG/6ML/30PKG (RAX; Thermo Scientific, cat. no. BQ0008Q1612)
- Parafilm® M (Sigma, cat. no. P7793-1EA)
- EURO DPC MicroMix 5 Shaker (Conquer Scientific, cat. no. 5957)
- µQuant Microplate Spectrophotometer (Biotek Instruments)
- Ultrasonic homogenizer (Branson, cat. no. 101-063-196R)
- Centrifuge 5415 D (Eppendorf, cat. no. 2262140-8)
- Innova 4000 incubator shaker (New Brunswick Scientific, cat. no. 1510611)
- Revco freezers (ultima II; Thermo Fisher Scientific, cat. no. 2586-9SI-V)
- Lab Dancer Mini Vortexer (VWR International, cat. no. 10153)
- Thermo Savant SpeedVac SPD121P Centrifugal Evaporator (Thermo Fisher Scientific, cat. no. SPD121PP1-115)
- Incubator set to 65 °C (Fisher Scientific, cat. no. 15-103-0514)
- GE Whatman 10360005 Universal Indicator Test Paper, Strips, 0 to 14 pH Range (pH strips; GE Whatman, cat. no. 10360005)
- NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, cat. no. ND-2000)
- Freezer set to -80 °C (Forma Scientific)
- Mass spectrometer: Orbitrap Fusion Tribrid mass spectrometer equipped with the EASY 1200 system (Thermo Fisher Scientific)

Software

- SEQUEST in Proteome Discoverer™ Software, v2.2
- MSConvertGUI (64-bit) in ProteoWizard
- Extract_oxonium_ion_spectra_intensity.exe script (download at <https://github.com/appleTree2017/Extraction-of-O-linked-glycopeptides-EXoO->)

REAGENT SETUP:

Tris-HCl (200mM, pH 7.4 or 8.0) Add 10 ml of 1M Tris-HCl to 40 ml of DI water.

Urea (8M)/Tris-HCl (200mM, pH 8.0) Fill a 15 ml BD Falcon tube with 4.8 g of urea powder. Add Tris-HCl (200 mM, pH 8.0) to the tube containing urea to 10 ml mark. Properly dissolve urea. Prepare the solution fresh.

DTT (1 M solution) Dissolve 7.7125 mg of DTT in 50 μ l DI water for a concentration of 1 M. Prepare the solution fresh.

IAA (500 mM solution) Dissolve 9.24 mg of IAA in 50 μ l of DI water. Prepare the solution fresh.

TFA (50%, vol/vol) in DI water Add 1ml of TFA in 1 ml DI water.

TFA (0.1%, vol/vol) in DI water Add 1 ml of TFA (50%) in 500 ml of DI water.

ACN (60%)/TFA (0.1%) Add 60 ml of ACN and 200 μ l of TFA (50%) in 39.8 ml of DI water.

ACN (50%)/TFA (0.1%) Add 50 ml of ACN and 200 μ l of TFA (50%) in 49.8 ml of DI water.

ACN (95%)/TFA (1%) Add 95 ml of ACN and 2 ml of TFA (50%) in 3 ml of DI water.

O-methylisourea solution (0.6 M solution) Dissolve 73.872 mg of O-Methylisourea hemisulfate salt in 1 ml of DI water. Prepare the solution fresh.

Guanination solution Mix equal volumes of NH₄OH, TFA (0.1%), and O-methylisourea solution (0.6 M). Check pH is 10.5 using pH strips. Prepare the solution fresh.

Sodium cyanoborohydride (1 M solution) Dissolve 62.84 mg of NaCNBH₃ in 1 ml DI water. The solution can be used for up to two days.

Triethylammonium acetate (100 mM solution) Add 5 ml triethylammonium acetate buffer (1 M pH 7.0) in 45 ml DI water. The buffer can be kept at 4°C.

NaCl (1.5 M) Add 75 ml of 5 M NaCl in 175 ml of DI water.

EQUIPMENT SETUP

LC setup

The given setting is for an Easy-nLC 1200 System (Thermo Scientific)

Column	In-house packed 20 cm \times 75 μ m diameter C18 column (1.9 μ m Reprosil-Pur C18-AQ beads). The column is heated to 50°C using a column heater (Phoenix-ST).
Mobile Phases	A: 3% ACN/0.1% formic acid in HPLC water B: 90% ACN/0.1% formic acid in HPLC water
Flow Rate (μ l per min)	2
Sample Injection Volume (μ l)	4
Gradient Program:	

Time (min)	% Solvent B
0	2
1	6
85	30
94	60
95	90
100	90
101	50
110	50

MS setup

The given setting is for Orbitrap Fusion Lumos.

Parameter	Orbitrap Fusion Lumos
Ion Source Type	NSI
Ionization mode	Positive
Electrospray voltage	1.8 kV
Internal Cal Positive	445.12003 m/z
MS1 Scan	
MS1 Resolution	60k
MS1 Scan Range (m/z)	350–1800
Detection Type	Orbitrap
MS1 Maximum Injection Time (ms)	50
MS1 AGC Target	4×10^5
RF Lens (%)	30
MIPS Mode	Peptide
Polarity	Positive
Data Dependent Mode	Cycle Time
Data Type	Profile
MS2 Scan	
Isolation Mode	Quadrupole
Orbitrap Resolution	50 k
MS2 Scan Range Mode	Auto Normal
First Mass (m/z)	110
Activation Type	HCD
Detector Type	Orbitrap
MS2 Maximum Injection Time (ms)	105
MS2 AGC Target	2×10^5
Isolation Window (m/z)	0.7

Parameter	Orbitrap Fusion Lumos
HCD Energy	36
Dynamic Exclusion	45 s
Exclude Isotopes	True
Charge States	2–6
Intensity Threshold	5e4
Data Type	Centroid
DDA Method	3 s cycle time

PROCEDURE:

Protein extraction, denaturation, reduction, alkylation, and digestion TIMING ~24 h

1. Extraction and denaturation of proteins from tissue, cells, and serum
 - a. **Extraction and denaturation of proteins from tissue**
 - i. Place the tube containing the tissue on ice.
 - ii. Transfer the tissue to a glass slide and cut the tissue into pieces as small as possible using a scalpel.
 - iii. Transfer the small pieces of tissue to a 1.5 ml Eppendorf tubes using pipette tips and forceps.
 - iv. Transfer remaining tissue pieces from the glass slide to the 1.5 ml Eppendorf tubes using 100 μ l of urea (8M)/Tris-HCl (200mM, pH 8.0). Repeat this step to recover the remaining tissue sample.
 - v. Add urea (8M)/Tris-HCl (200mM, pH 8.0) to reach a ratio of buffer/tissue volume of 4:1.
 - vi. Sonicate the sample at least three rounds using ultrasonic homogenizer with duty cycle at 20% and output control at 2 for 2 min. Between rounds, cool the sample on ice for at least 1 min.
 - vii. Centrifuge the lysate at 16,000g for 10 min, 4°C.
 - viii. Collect the supernatant containing the extracted proteins.
 - b. **Extraction and denaturation of proteins from cells**
 - i. Harvest cells from culture.
For adherent cell culture, remove the culture medium. Scrape cells from the culture plate and transfer the cells to a BD Falcon tube.

CRITICAL STEP Harvesting adherent cells using trypsin-EDTA solution will significantly decrease recovery of glycoproteins since many glycoproteins on the cell surface are trimmed off by trypsin digestion.

For suspension cell culture, transfer cells to a BD Falcon tube.

- ii. Centrifuge at 400g for 10 min, 4°C.
- iii. Remove supernatant.
- iv. Thoroughly resuspend cells in 10 ml PBS.
- v. Repeat the above steps (ii-iv) for a total of three times.
- vi. Add urea (8M)/Tris-HCl (200mM, pH 8.0) to reach a ratio of buffer/cell volume of 4:1.
- vii. Sonicate the sample at least three rounds using ultrasonic homogenizer with duty cycle at 20% and output control at 2 for 2 min. Between rounds, cool the sample on ice for at least 1 min.
- viii. Centrifuge the lysate at 16,000g for 10 mins, 4°C.
- ix. Collect the supernatant containing the extracted proteins.

c. Denaturation of protein from serum

- i. Add 45 µl of urea (8M)/Tris-HCl (200mM, pH 8.0) to 10 µl of serum.

PAUSE POINT The solution containing extracted proteins can be left overnight at 4°C or frozen for up to a month at -20°C or at -80°C for long-term storage.

2. Measurement of protein concentration using BCA

- i. Prepare reference series (0, 0.0625, 0.125, 0.25, 0.5, 1, and 2 mg/ml) using BSA protein standard. Transfer 25 µl of reference series to a 96-well plate.
- ii. Dilute protein samples in 1× PBS. The sample can be diluted 10-fold followed by 2-fold or 10-fold series dilutions.
- iii. Add 25 µl of diluted protein sample to the 96-well plate.
- iv. Mix BCA reagent B with A at a ratio of 1:50. Make 200 µl for each well.
- v. Add 200 µl of mixed BCA reagent B and A to the 96-well plates.
- vi. Mix the samples using a DPC shaker for 1 min.
- vii. Incubate the 96-well plate for 30 min, 37°C.
- viii. Measure the protein concentration using µQuant Microplate Spectrophotometer.

- ix. Calculate the amount of proteins according to their concentration and volume.

PAUSE POINT The solution containing extracted proteins can be left overnight at 4°C or frozen for up to a month at –20°C or at –80°C for long-term storage.

3. Reduction: take the desired volume of protein sample and add 1M DTT to a final concentration of 5 mM, mix, and incubate for 1h, 37°C. For instance, add 5 µl of 1M DTT to 1 ml of sample.
4. Alkylation: add 500 mM IAA to a final concentration of 10 mM, mix, and incubate for 40 min in the dark, room temperature. For instance, add 20 µl of 500 mM IAA to 1 ml of sample.
5. Digestion: dilute the sample at least five times using Tris-HCl (200 mM, pH 8.0) and add trypsin to an enzyme/protein ratio of 1:40.
6. Incubate for 16 h or overnight, room temperature.

Note: it is not necessary to incubate at 37°C. Higher temperature may cause carbamylation due to the use of urea.

PAUSE POINT The solution after trypsin digestion can be left overnight at 4°C or frozen for up to a month at –20°C or at –80°C for long-term storage.

Sequential guanidination and desalting on the same C18 cartridge TIMING ~1 h

7. Check or adjust incubator temperature to 65 °C.
8. Add TFA (50%) to samples to a final concentration of 1%. For instance, add 20 µl of TFA (50%) to 1 ml of the sample after trypsin digestion. Mix and check pH < 3 using pH strips.

CRITICAL STEP pH is vital for C18 to retain peptides.

? TROUBLESHOOTING

9. Centrifuge the samples at 16,000g for 10 min.
10. Transfer supernatant to a new tube and discard the pellet.
11. Condition C18 cartridge with three times of ACN (60%)/TFA (0.1%), three times of TFA (0.1%).

CRITICAL STEP Do not dry C18 cartridge.

Note:

- The filter membrane on the top of C18 bedding material is suitable to keep the cartridge wet for 30 min. Load conditioned C18 cartridge with TFA (0.1%) and seal the top of C18 cartridge using tape to prevent dryness of the cartridge when the samples are not ready.

- The capacity of the C18 cartridge is 1–3% of bedding material. For instance, C18 cartridge with 100 mg of C18 bedding material is suitable for desalting peptides from 1–3 mg of proteins.
 - Suction or push force using air bulb can be used to force solutions to pass C18 cartridge faster to speed up conditioning of the cartridge. However, pay attention to the level of solution to avoid dryness of C18 cartridge.
12. Load the sample and allow the sample to flow through the C18 cartridge by gravity.
 13. Wash C18 cartridge three times using TFA (0.1%).
 14. Wash C18 cartridge three times using guanidination solution.
 15. Apply 1-bed volume of guanidination solution to the C18 cartridge and seal C18 cartridge using tape for the top and plastic blocker from snap-cap spin column kit for the bottom.
 16. Incubate the C18 cartridge in the incubator for 20 min, 65 °C.
CRITICAL STEP Open and close the door of the incubator may slightly lower the temperature. Keep the temperature above 60°C. Incubate for 20 min only. Longer or shorter incubation time may affect the number of identifications.
 17. Transfer the C18 cartridge to 4°C fridge for 5 min to cool down the cartridge.
 18. Wash C18 cartridge three times using TFA (0.1%).
 19. Elute peptide using ACN (60%)/TFA (0.1%).
Note: use 200, 400, or 2000 µl of ACN (60%)/TFA (0.1%) for C18 cartridge with 50, 100, and 500 mg C18 bedding material, respectively.
 20. Determine the concentration of peptides using Nanodrop. Calculate the amount of peptides according to its concentration and volume.
Note: Recovery rate of peptides varies among samples. A reasonable peptide amount may range from 40–90% of starting proteins.

RAX enrichment of glycopeptides TIMING ~1 h

21. adjust ACN (60%)/TFA (0.1%) as C18 elute to ACN (95%)/TFA (1%) by adding ACN and TFA (50%). For instance, add 2800 µl of ACN and 64 µl of TFA (50%) to 400 µl of C18 elute containing ACN (60%)/TFA (0.1%).
22. Determine the amount of bedding material of SAX cartridge. RAX cartridge with 30 mg of bedding material is suitable for 1 mg or less peptide sample.
23. Condition the RAX cartridge using three times of ACN, three times of triethylammonium acetate (100 mM, pH 7.0), three times of DI water, and three times of ACN (95%)/TFA (1%).

CRITICAL STEP Do not allow the RAX cartridge to run dry.

Note: the membrane on the top of bedding material may keep the cartridge wet for 2 min. Load conditioned RAX cartridge with ACN (95%)/TFA (1%) and seal the top of the cartridge using tape to prevent dryness of the cartridge when the samples are not ready.

24. Load the RAX cartridge with samples (step 21) and allow the sample to flow through the cartridge by gravity.
25. Wash the cartridge with ACN (95%)/TFA (1%) for three times.
26. Elute glycopeptides using 1-bed volume of ACN (50%)/TFA (0.1%).
27. Determine the concentration of glycopeptides using Nanodrop. Calculate the amount of glycopeptides according to its concentration and volume.

Note: the amount of glycopeptides will be used to calculate resin (step 30) and enzymes (step 59).

Conjugation and release of site-containing O-GalNAc glycopeptides TIMING ~24 h

28. Add 5-fold volume of 10× PBS pH 7.4 to RAX elute in ACN (50%)/TFA (0.1%).
? TROUBLESHOOTING
 29. Determine pH = ~7.0 using pH strips.
? TROUBLESHOOTING
 30. Transfer AminoLink plus coupling resin to a new tube. One μ l of resin is suitable for binding 1 μ g of peptides determined at step 27.
 31. Wash the resin: add 1 ml of DI water to the resin, vortex for 10 sec, centrifuge the resin at 2,000g for 30 sec, and discard the supernatant.
 32. Repeat step 31 for a total of three times.
 33. Transfer the resin to the peptide sample (step 29).
 34. Add NaCNBH₃ (1 M) to a final concentration of 50 mM and incubate at least 4 h or overnight with gentle rotation, RT. For instance, add 20 μ l of NaCNBH₃ (1 M) to 1 ml of the sample containing the mixture of glycopeptides and resin.
? TROUBLESHOOTING
- PAUSE POINT** The solution containing the mixture of glycopeptides and resin can be stored for up to two days at 4°C. Do not freeze.
35. Collect resin by transferring the sample to snap-cap spin column.
 36. Centrifuge at 1,000g for 10 sec. Discard flow through. Apply multiple time if the volume of solution large than that of the spin column.
 37. Block bottom of the spin column using plastic blocker from the spin column kit.
 38. Add 650 μ l of Tris-HCl (1 M, pH 7.5) to spin column containing resin.

39. Add 32.5 μ l of NaCNBH₃ (1 M) to spin column to block unused aldehyde group on the resin.
40. Incubate for 30 min with gentle rotation, RT.
PAUSE POINT The solution containing resin can be stored for up to two days at 4°C. Do not freeze.
41. Remove plastic blocker and centrifuge the spin column at 2,000g for 10 sec. Discard the flow-through.
42. Block spin column with plastic blocker and add 650 μ l of ACN (60%)/TFA (0.1%) to spin column containing the resin and vortex for 1 min.
43. Remove plastic blocker and centrifuge the spin column at 2,000g for 10 sec. Discard the flow-through.
44. Repeat steps 42–43 for at least three times.
45. Add 650 μ l of NaCl (1.5 M) to spin column containing the resin and vortex for 1 min.
46. Centrifuge the spin column at 2,000g for 10 sec. Discard the flow-through.
47. Repeat steps 45–46 for at least three times.
48. Add 650 μ l of Tris-HCl (200 mM, pH 7.4) to spin column containing the resin and vortex for 1 min.
49. Centrifuge the spin column at 2,000g for 10 sec. Discard the flow-through.
50. Repeat steps 48–49 for at least three times.
51. Transfer resin to a new 1.5 ml Eppendorf tube by add 400 μ l of Tris-HCl (200 mM, pH 7.4) to the spin column and transferring the solution containing the resin to the 1.5 ml tube.
52. Repeat step 51 once, combine the two collections of resin.
53. Centrifuge 2,000g for 30 sec.
54. Place the tube upright for 5 min to allow the resin to settle.
55. Remove and discard the supernatant.
56. Add 50 μ l of Tris-HCl (200 mM, pH 7.4) to the resin.
57. Add 50 μ l of DI water to a tube containing SialEXO. The concentration is 40 unit/ μ l.
58. Add 50 μ l of DI water to a tube containing OperATOR. The concentration is 40 unit/ μ l.
59. Add SialEXO and OperATOR to the tube (step 56) and mix by pipetting.
CRITICAL STEP Do not vortex. The resin may stick to the wall of the tube if vortex.

Note: one unit of SialEXO or OpeRATOR is suitable for 1 µg of glycopeptides determined at step 27.

60. Incubate 16 h or overnight, 37°C.

Preparation of site-containing O-GalNAc glycopeptides for LC-MS/MS TIMING ~6 h

61. Add 400 µl of Tris-HCl (200 mM, pH 7.4) to resin and vortex for 1 min.
62. Centrifuge at 2,000g for 30 sec.
63. Collect supernatant in a new 1.5 ml Eppendorf tube.

Note: the supernatant contains the site-containing O-GalNAc glycopeptides.

64. Repeat step 61–63 once and combine the two collections of supernatant in the 1.5 ml Eppendorf tube. Discard the resin.
65. Centrifuge the sample at 2,000g for 1 min.
66. Place the sample upright for 5 min to allow the resin to settle.
67. Transfer the supernatant to a new 1.5 ml Eppendorf tube.

CRITICAL STEP Avoid taking resin at the bottom of the tube.

Note: steps 65–67 removes the remaining resin in the sample.

PAUSE POINT The solution containing site-containing O-GalNAc glycopeptides can be left overnight at 4°C or frozen for up to a month at –20°C or at –80°C for long-term storage.

68. Add 16 µl of TFA (50%) and check the pH < 3 using pH strips.

CRITICAL STEP pH < 3 in the solution is important for C18 to retain peptides.

? TROUBLESHOOTING

69. Condition C18 cartridge, load sample, and wash the cartridge as described at steps 11–13.
70. Elute site-containing O-GalNAc glycopeptides in 1-bed volume of ACN (60%)/TFA (0.1%).
71. Determine the concentration of glycopeptides using Nanodrop. Calculate the amount of glycopeptides according to its concentration and volume.

Note: the amount of glycopeptides will be used for calculation at step 73.

72. Dry the sample in a speed-vac.

PAUSE POINT The dried glycopeptides can be stored at –20°C or –80°C for long-term storage.

73. Re-suspend the samples in 100 µl TFA (0.1%) for every 100 µg of peptides determined at step 71.
74. Centrifuge at 16,000g for 10 min.

- 75. Transfer supernatant to a new 1.5 ml Eppendorf tube. Discard undissolved particles.
- 76. Determine the concentration of glycopeptides using Nanodrop. Calculate the amount of glycopeptides according to its concentration and volume.

PAUSE POINT The solution contains glycopeptides can be stored at -20°C or -80°C for long-term storage.

LC-MS/MS analysis TIMING 2 h per sample

- 77. Before LC-MS/MS analysis, centrifuge sample at 16,000g for 10 min.
 CRITICAL STEP This step removes any particular matter that may clog the LC column.
- 78. Take the required volume of sample and dilute the sample with TFA (0.1%) to 1 μg per injection. If injection volume is 6 μl , prepare 1 μg peptide per 6 μl . For instance, if the concentration of the sample is 1 $\mu\text{g}/\mu\text{l}$, take 1 μl sample and add 5 μl of TFA (0.1%) to make a total volume of 6 μl for an LC-MS/MS run.
- 79. Setup LC-MS/MS system as described in the Equipment Setup section and analyze the sample.

? TROUBLESHOOTING

Database search TIMING variable; hours to 1 d per LC-MS/MS file, depending on sample complexity and computer hardware

- 80. SEQUEST in Proteome Discoverer v2.2 is used to search the site-containing O-GalNAc glycopeptides with the following parameters:

Processing Step:

- i. Spectrum Files RC node with the following settings:

Search settings	
Protein Database	your database in FASTA format
Enzyme Name	Trypsin (Full)
Precursor Mass Tolerance	20 ppm
Fragment Mass Tolerance	0.5 Da
Dynamic Modification	Oxidation (+15.995 Da at methionine)
Static Modification	Carbamidomethyl (+57.021 Da at cysteine)
Static Modification	Guanidinylyl (+42.022 Da at lysine)

- ii. Spectrum Selector node – default
- iii. SEQUEST HT node with following settings:

Input Data	
Protein Database	your database in FASTA format
Enzyme Name	Trypsin (semi)
Max. Missed Cleavage	2
Min. Peptide length	6
Max. Peptide length	50
Tolerances	
Precursor Mass Tolerance	10 ppm
Fragment Mass Tolerance	0.06 Da
Dynamic Modifications	
Oxidation	+15.995 Da at methionine
GuanidinyI	+42.022 Da at lysine
HexNAc	+203.079 Da at serine and threonine
Hex(1)HexNAc(1)	+365.132 Da at serine and threonine
Static Modification	
Carbamidomethyl	+57.021 Da at cysteine

iv. Percolator node - default

Consensus Step: contains the following nodes with default setting: MSF Files, PSM Grouper, Peptide Validator, Peptide and Protein Filter, Protein Scorer, Protein FDR Validator, and Protein Grouping.

Data filtering TIMING ~2 h

81. After the search, go to PSMs tag and apply the following filters:
 - i. Modification: contains HexNAc
 - ii. Rank: is equal to 1
 - iii. XCorr: is greater than or equal to 1
82. Add the following column headers to the PSMs tag:
 - i. Sequence
 - ii. Protein Descriptions
 - iii. Master Protein Descriptions
83. Export data: click PSMs tag followed by clicking File\Export\To Microsoft Excel. Select directory path to save the file. Pick Level 1: PSMs. Click export to save the file.
84. Open the exported Excel file. Sort column header Sequence. Keep peptide sequence starting with Ser and Thr and remove other peptides.
85. Open MSConvertGUI (64-bit) after installation of ProteoWizard. Convert the MS raw file to mzML format after adding filters: Peak Picking.

- 86.** Download [extract_oxonium_ion_spectra_intensity.exe](#) file and follow the instruction in [extract_oxonium_ion_spectra_intensity.pdf](#) to run the script. Briefly, double click the .exe, input directory path of the mzML and press enter.

Note: the script will extract MS/MS scan number containing oxonium ions at 204 m/z being mandatory with at least one another oxonium ion. List of oxonium ions is in Table 1.

- 87.** Keep PSMs identified in both steps 84 and 86. The data is ready for reporting and downstream bioinformatics analysis.

? TROUBLESHOOTING

Troubleshooting advice can be found in Table 2.

TIMING

Steps 1–6, protein extraction, denaturation, reduction, alkylation, and digestion: ~24 h

Steps 7–20, sequential guanidination and desalting on the same C18 cartridge: ~1 h

Steps 21–27, RAX enrichment of glycopeptides: ~1 h

Steps 28–60, conjugation and release of site-containing O-GalNAc glycopeptides: ~24 h

Step 61–76, preparation of site-containing O-GalNAc glycopeptides: ~6 h

Steps 77–79, LC-MS/MS analysis: 2 h per sample

Step 80, data search: variable; hours to 1 d per LC-MS/MS file, depending on sample complexity and computer hardware.

Steps 81–87, data filtering and reporting: ~2 h

ANTICIPATED RESULTS

The protocol is used to achieve large-scale identification of O-GalNAcylation sites in the complex samples. Mapping of O-GalNAcylation sites is difficult that is typically achieved using ETD fragmentation technique. ETD preferentially fragments peptide backbones while leaves O-GalNAc glycans intact on the peptides to facilitate site localization. ETD, however, suffers limitations so that its capability to identify large-scale O-GalNAcylation sites is not optimal³⁶. Also, the mucin domain in glycoproteins is resistant to digestion by common proteases that hamper the study of O-GalNAcylation on mucin domains. This protocol describes an effective approach to surmount these issues by using an O-GalNAc glycan dependent endoprotease OpeRATOR. OpeRATOR cleaves N-termini of O-glycan-occupied Ser and Thr so that the first Ser or Thr residues in the identified glycopeptides are the glycosylation sites. O-GalNAc glycans are still attached to the released glycopeptides that yield oxonium ions in MS/MS spectra providing additional confidence for identification. The use of solid phase enables optimal removal of contaminants to facilitate detection of

glycopeptides since the presence of peptides would suppress the signal of glycopeptides during LC-MS/MS analysis³⁷.

Example of the protocol

The protocol has been used to identify over 3,000 O-GalNAcylation sites from over 1,000 glycoproteins in human kidney tissue, serum, and T cells¹⁶. Precisely, 1,781, 732 and 1,295 O-GalNAcylation sites are mapped in kidney tissue, serum and T cells, respectively¹⁶. Approximately 69% of total PSM are assigned with Hex(1)HexNAc(1) glycan composition, most likely Gal-GalNAc glycan structure. Four MS/MS spectra are showed to illustrate HCD-MS2 assignments and data interpretation (Fig. 2). In all assigned MS/MS spectra, oxonium ions are seen at the left side of the spectra (Fig. 2). The MS/MS spectra of glycopeptides should contain oxonium ion at 204 m/z and at least another oxonium ion in table 1 to be kept for the assignment. For peptide with short or medium length, a peptide ion is commonly seen on the rightmost side of the MS/MS spectra (Fig. 2A, C, and D). Peptide fragment b- and y- ions are observed in the middle of the spectra (Fig. 2). For long peptide, the peptide ion may not be seen on the rightmost side of the spectrum (Fig. 2B). Instead, the spectrum is rich with peptide b- and y- fragment ions (Fig. 2B). The fragmentation pattern is relatively conserved for the assigned glycopeptides. Regarding the glycan composition, the glycopeptides can be modified by Hex(1)HexNAc(1) or HexNAc(1) that can be predicted to be a single glycan attached to the glycopeptides (Fig. 2A, B, and C). However, when the glycan composition is identified as Hex(2)HexNAc(2) or large, the attached glycan may consist of a single Hex(2)HexNAc(2) or two Hex(1)HexNAc(1) (Fig. 2D). After the identification of site-specific O-GalNAc glycoproteome, protein accession number can be analyzed using GO analysis in DAVID Bioinformatics³⁸. An integral component of membrane, ER, Golgi apparatus, cell surface, and extracellular space are typically associated with the identified proteins¹⁶. The GO analysis gives some validation to the identification of O-GalNAc glycoproteome.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, National Cancer Institute, the Early Detection Research Network (EDRN, U01CA152813), the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA210985), and amfAR, The Foundation for AIDS Research on Bringing Bioengineers to Cure HIV (Grant amfAR 109551-61-RGRL).

Data availability

The data is available in PRIDE partner repository³⁹ with the dataset identifier: project accession: PXD009476

Reference

1. Gahmberg CG & Tolvanen M Why mammalian cell surface proteins are glycoproteins. *Trends in biochemical sciences* 21, 308–311 (1996). [PubMed: 8772385]
2. Jang JH & Hanash S Profiling of the cell surface proteome. *Proteomics* 3, 1947–1954, doi:10.1002/pmic.200300563 (2003). [PubMed: 14625857]
3. Kovtun YV & Goldmacher VS Cell killing by antibody-drug conjugates. *Cancer letters* 255, 232–240, doi:10.1016/j.canlet.2007.04.010 (2007). [PubMed: 17553616]

4. Selvaraj P et al. The T lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3. *Nature* 326, 400–403, doi:10.1038/326400a0 (1987). [PubMed: 2951597]
5. Orentas RJ et al. Identification of cell surface proteins as potential immunotherapy targets in 12 pediatric cancers. *Frontiers in oncology* 2, 194, doi:10.3389/fonc.2012.00194 (2012). [PubMed: 23251904]
6. Posey AD Jr. et al. Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* 44, 1444–1454, doi:10.1016/j.immuni.2016.05.014 (2016). [PubMed: 27332733]
7. Westerlund B & Korhonen TK Bacterial proteins binding to the mammalian extracellular matrix. *Molecular microbiology* 9, 687–694, doi:10.1111/j.1365-2958.1993.tb01729.x (1993). [PubMed: 7901732]
8. Maddon PJ et al. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47, 333–348, doi:10.1016/0092-8674(86)90590-8 (1986). [PubMed: 3094962]
9. Kennedy S Proteomic profiling from human samples: the body fluid alternative. *Toxicology letters* 120, 379–384 (2001). [PubMed: 11323197]
10. Brockhausen I & Stanley P O-GalNAc Glycans in *Essentials of Glycobiology*. 3rd edn (eds Varki A *et al.*) (2015).
11. Tarelli E, Smith AC, Hendry BM, Challacombe SJ & Pouria S Human serum IgA1 is substituted with up to six O-glycans as shown by matrix assisted laser desorption ionisation time-of-flight mass spectrometry. *Carbohydrate research* 339, 2329–2335, doi:10.1016/j.carres.2004.07.011 (2004). [PubMed: 15337464]
12. Arnold JN et al. The glycosylation of human serum IgD and IgE and the accessibility of identified oligomannose structures for interaction with mannan-binding lectin. *Journal of immunology* 173, 6831–6840, doi:10.4049/jimmunol.173.11.6831 (2004).
13. Sasaki H, Bothner B, Dell A & Fukuda M Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. *The Journal of biological chemistry* 262, 12059–12076 (1987). [PubMed: 3624248]
14. Yang W et al. Glycoform Analysis of Recombinant and Human Immunodeficiency Virus Envelope Protein gp120 via Higher Energy Collisional Dissociation and Spectral-Aligning Strategy. *Analytical chemistry* (2014).
15. Hang HC & Bertozzi CR The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 13, 5021–5034, doi:10.1016/j.bmc.2005.04.085 (2005). [PubMed: 16005634]
16. Yang W, Ao M, Hu Y, Li QK & Zhang H Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). *Molecular systems biology* 14, e8486 (2018). [PubMed: 30459171]
17. Steentoft C et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8, 977–982, doi:10.1038/nmeth.1731 (2011). [PubMed: 21983924]
18. Woo CM, Iavarone AT, Spicciarich DR, Palaniappan KK & Bertozzi CR Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* 12, 561–567, doi:10.1038/nmeth.3366 (2015). [PubMed: 25894945]
19. Nilsson J et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods* 6, 809–811, doi:10.1038/nmeth.1392 (2009). [PubMed: 19838169]
20. Medzihradsky KF, Kaasik K & Chalkley RJ Tissue-Specific Glycosylation at the Glycopeptide Level. *Mol Cell Proteomics* 14, 2103–2110, doi:10.1074/mcp.M115.050393 (2015). [PubMed: 25995273]
21. Hoffmann M, Marx K, Reichl U, Wuhler M & Rapp E Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol Cell Proteomics* 15, 624–641, doi:10.1074/mcp.M115.053546 (2016). [PubMed: 26598643]
22. Kufe DW Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer* 9, 874–885, doi:10.1038/nrc2761 (2009). [PubMed: 19935676]
23. Tabak LA The role of mucin-type O-glycans in eukaryotic development. *Semin Cell Dev Biol* 21, 616–621, doi:10.1016/j.semcdb.2010.02.001 (2010). [PubMed: 20144722]

24. Yang S et al. Deciphering Protein O-Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment. *Analytical chemistry* 90, 8261–8269, doi:10.1021/acs.analchem.8b01834 (2018). [PubMed: 29936827]
25. Sun S et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nature biotechnology* 34, 84 (2016).
26. Zhang H, Li XJ, Martin DB & Aebersold R Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21, 660–666, doi:10.1038/nbt827 (2003). [PubMed: 12754519]
27. Yang S, Hu YW, Sokoll L & Zhang H Simultaneous quantification of N- and O-glycans using a solid-phase method. *Nature Protocols* 12, 1229–1244, doi:10.1038/nprot.2017.034 (2017). [PubMed: 28518173]
28. Huang J et al. Highly Efficient Release of Glycopeptides from Hydrazide Beads by Hydroxylamine Assisted PNGase F Deglycosylation for N-Glycoproteome Analysis. *Analytical chemistry* 87, 10199–10204, doi:10.1021/acs.analchem.5b02669 (2015). [PubMed: 26399494]
29. Yang W et al. Comparison of enrichment methods for intact N-and O-linked glycopeptides using strong anion exchange and hydrophilic interaction liquid chromatography. *Analytical chemistry* 89, 11193–11197 (2017). [PubMed: 29016103]
30. Zhu Z, Su X, Clark DF, Go EP & Desaire H Characterizing O-linked glycopeptides by electron transfer dissociation: fragmentation rules and applications in data analysis. *Analytical chemistry* 85, 8403–8411, doi:10.1021/ac401814h (2013). [PubMed: 23909558]
31. Good DM, Wirtala M, McAlister GC & Coon JJ Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 6, 1942–1951, doi:10.1074/mcp.M700073-MCP200 (2007). [PubMed: 17673454]
32. Darula Z, Sherman J & Medzihradzky KF How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol Cell Proteomics* 11, O111 016774, doi:10.1074/mcp.O111.016774 (2012).
33. Mulagapati S, Koppolu V & Raju TS Decoding of O-Linked Glycosylation by Mass Spectrometry. *Biochemistry* 56, 1218–1226, doi:10.1021/acs.biochem.6b01244 (2017). [PubMed: 28196325]
34. Jensen PH, Kolarich D & Packer NH Mucin-type O-glycosylation--putting the pieces together. *The FEBS journal* 277, 81–94, doi:10.1111/j.1742-4658.2009.07429.x (2010). [PubMed: 19919547]
35. Malaker SA et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proceedings of the National Academy of Sciences of the United States of America* 116, 7278–7287, doi:10.1073/pnas.1813020116 (2019). [PubMed: 30910957]
36. Mechref Y Use of CID/ETD mass spectrometry to analyze glycopeptides. *Current protocols in protein science* Chapter 12, Unit 12 11 11–11, doi:10.1002/0471140864.ps1211s68 (2012). [PubMed: 22470127]
37. Sutton CW, O'Neill JA & Cottrell JS Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser desorption mass spectrometry. *Analytical biochemistry* 218, 34–46, doi:10.1006/abio.1994.1138 (1994). [PubMed: 8053566]
38. Dennis G Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, P3 (2003). [PubMed: 12734009]
39. Vizcaino JA et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44, D447–456, doi:10.1093/nar/gkv1145 (2016). [PubMed: 26527722]

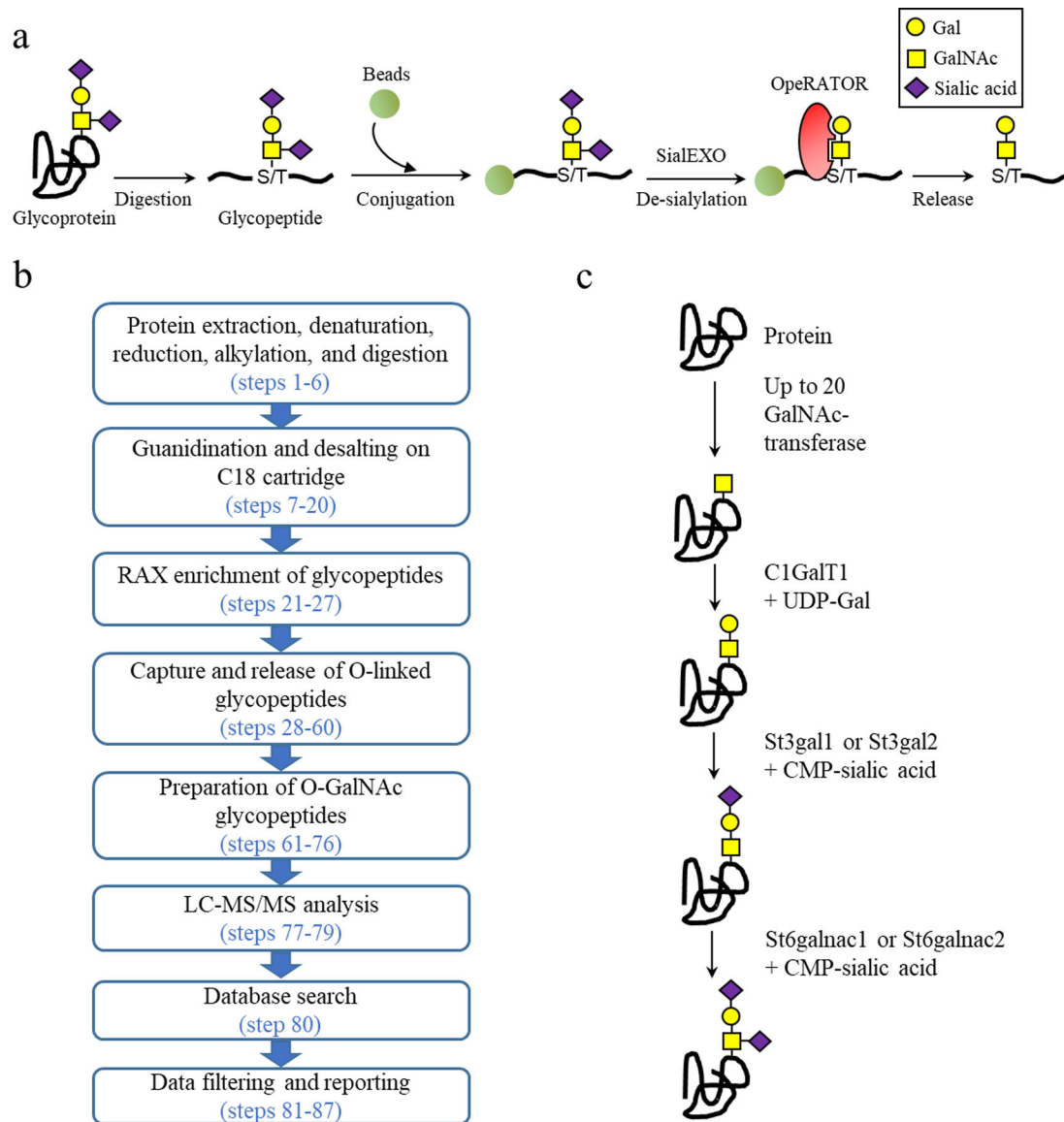


Figure 1 |. Workflow diagram of EXoO protocol and O-GalNAcylation pathway for Gal(1)GalNAc(1)Sia(2). (a) Schematic representation of EXoO method. Image is adapted and from ref.16. (b) Steps of EXoO protocol. (c) O-GalNAcylation pathway for Gal(1)GalNAc(1)Sia(2).

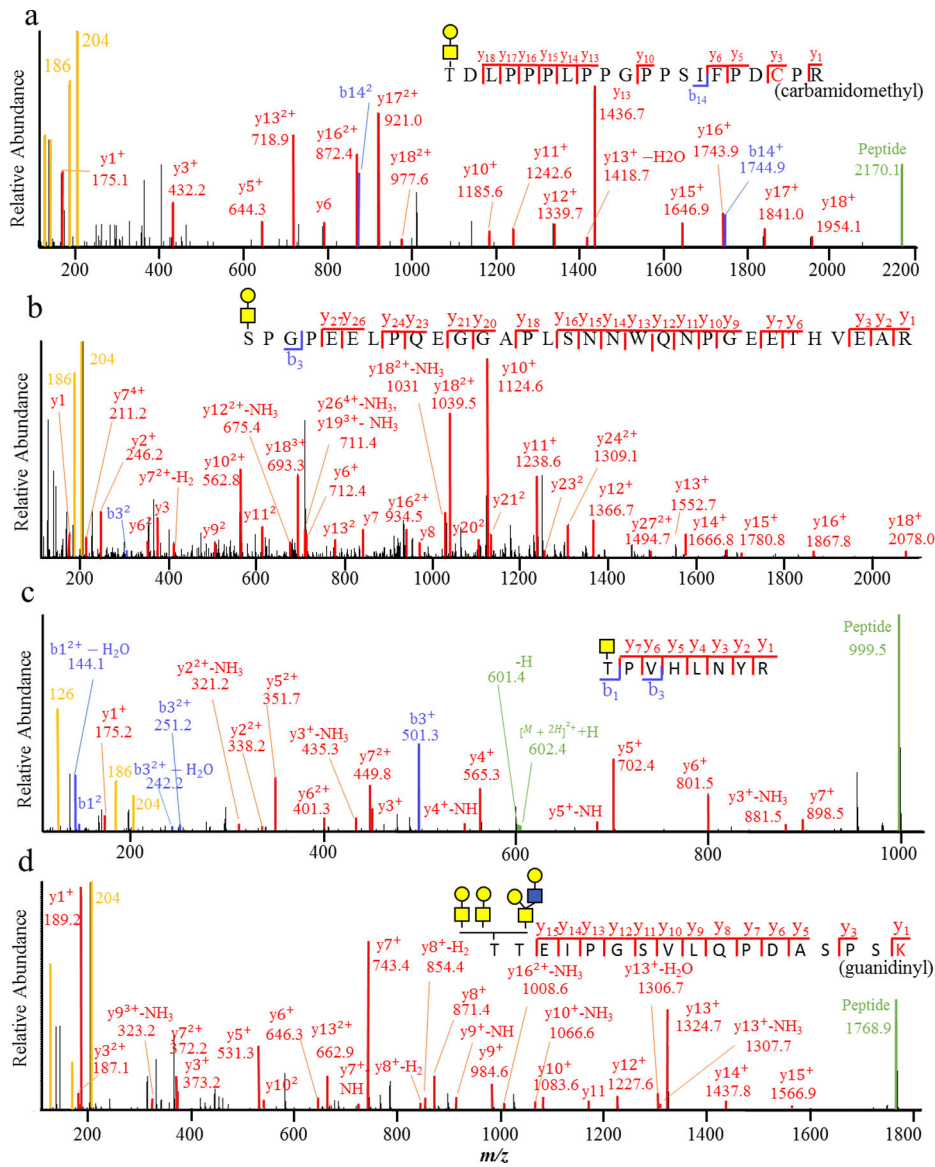


Figure 2 | MS/MS spectra of site-specific O-GalNAc glycopeptides. The glycopeptides are generated using the protocol. O-GalNAc glycosylation sites are at the N-termini of peptide sequences. Different glycan compositions are showed to illustrate the approach for defining site-specific O-GalNAc glycans. (a) MS/MS spectra of a glycopeptide with Hex(1)HexNAc(1). (b) MS/MS spectra of a short glycopeptide with a HexNAc. (c) MS/MS spectra of a long glycopeptide with Hex(1)HexNAc(1). (d) MS/MS spectra of a glycopeptide with Hex(2)HexNAc(2).

Table 1 |

List of oxonium ions in the MS/MS spectra of glycopeptides

Oxonium ion (m/z)	Glycan
126.055	[HexNAc-C ₂ H ₆ O ₃] ⁺
138.055	[HexNAc-CH ₆ O ₃] ⁺
144.066	[HexNAc-C ₂ H ₄ O ₂] ⁺
168.0655	[HexNAc-2H ₂ O] ⁺
186.0761	[HexNAc-H ₂ O] ⁺
204.0866	[HexNAc] ⁺
274.0927	[Neu5Ac-H ₂ O] ⁺
292.1032	[Neu5Ac] ⁺
366.140	[HexHexNAc] ⁺

Table 2 |

Troubleshooting table.

Step	Problem	Possible reason	Solution
8	No or low peptide yield after C18 desalting	pH is not low enough for C18 to bind to peptides.	Check pH < 3 using pH strips. Add more TFA (50%) to sample if needed.
28	Low yield of site-containing O-GalNAc glycopeptides	The buffer contains amino groups.	Avoid using Tris-HCl buffer or buffers contain amino groups.
29	pH is not ~7.0 that may lead to low yield of site-containing O-GalNAc glycopeptides	Buffer capacity of 10× PBS is not sufficient to neutralize RAX elute.	Add more 10× PBS and determine the pH = ~7.0 using pH strips.
34	Leakage of the solution after adding NaCNBH3 leads to low yield of site-containing O-GalNAc glycopeptides	Air bubbles are seen in NaCNBH3 solution. The air bubble may generate air pressure inside the tube that results in leakage of the solution during incubation.	Seal the tube with layers of parafilm to prevent leakage. Use 1.5 ml Eppendorf tube or 15 ml BD Falcon tube. Avoid using 50 ml BD Falcon tube that is found to have a higher chance of leakage.
68	Same to step 8	Same to step 8	Same to step 8
79	LC-MS/MS issues	LC-MS/MS may have issues in LC or MS instrument.	Consult MS core facility or scientist with expertise in LC-MS/MS instrumentation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript