

1

2 **EXoO-Tn: Tag-n-Map the Tn Antigen in the Human Proteome**

3

4 *Running title: Tag-n-Map Tn in the Human Proteome*

5

6

7 *Weiming Yang*, Minghui Ao, Angellina Song, Yuanwei Xu, and Hui Zhang*

8

9

10

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

13

14

15 ***Corresponding Author**

16 **Address:** Department of Pathology, Johns Hopkins University School of Medicine, 400 North
17 Broadway, Room 4001A, Baltimore, Maryland, United States.

18 **E-mail address:** wyang21@jhmi.edu

19 **Abstract**

20 Tn antigen (Tn), a single N-acetylgalactosamine (GalNAc) monosaccharide attached to protein
21 Ser/Thr residues, is found on most solid tumors yet rarely detected in adult tissues, featuring it
22 one of the most distinctive signatures of cancers. Although it is prevalent in cancers, Tn-
23 glycosylation sites are not entirely clear owing to the lack of suitable technology. Knowing the
24 Tn-glycosylation sites will spur the development of new vaccines, diagnostics, and therapeutics
25 of cancers. Here, we report a novel technology named EXoO-Tn for large-scale mapping of Tn-
26 glycosylation sites. EXoO-Tn utilizes glycosyltransferase C1GalT1 and isotopically-labeled
27 UDP-Gal(¹³C₆) to tag and convert Tn to Gal(¹³C₆)-Tn, which has a unique mass being
28 distinguishable to other glycans. This exquisite Gal(¹³C₆)-Tn structure is recognized by
29 OpeRATOR that specifically cleaves N-termini of the Gal(¹³C₆)-Tn-occupied Ser/Thr residues to
30 yield site-containing glycopeptides. The use of EXoO-Tn mapped 947 Tn-glycosylation sites
31 from 480 glycoproteins in Jurkat cells. Given the importance of Tn in diseases, EXoO-Tn is
32 anticipated to have broad utility in clinical studies.

33

34 **Keywords:** Tn antigen, site-specific, O-GalNAc, glycoproteomics, cancer

35 **Introduction**

36 Over decades of biomedical investigations, it was found that one of the most distinctive
37 features of cancers is the expression of Tn antigen (Tn), which is an N-acetylgalactosamine
38 (GalNAc) attached to protein Ser/Thr residues via an O-linked glycosidic linkage¹. A variant of
39 Tn is STn, which has an addition of sialic acid monosaccharide¹. Tn establishes its nature as a
40 pan-carcinoma antigen by finding of its expression in 10-90% of the solid tumor including lung,
41 prostate, breast, colon, pancreas, gastric, stomach, ovary, cervix, bladder¹⁻³. In sharp contrast,
42 the expression of Tn in adult tissue is rare⁴, making it an attractive target for anti-cancer
43 applications. For instance, Slovin et al. report a Phase I clinical trial using a vaccine consisting
44 of synthetic Tn on a carrier protein for prostate cancer⁵. Studies explore the potential of Tn for
45 early diagnostics⁶⁻⁸ and prognostics of cancers⁹⁻¹¹. To treat cancers, Posey et al. report the
46 development of engineered CAR-T cells that target Tn on mucin protein MUC1 (MUC1-Tn) for
47 killing cancer cells¹². Also, a Phase I clinical trial using MUC1-Tn specific CAR-T cells started
48 for treating patients with head and neck cancer^{13,14}. Despite a noteworthy link between Tn and
49 cancers, the underlying mechanism causing the expression of Tn in cancers is not entirely clear.
50 It may involve glycosyltransferase C1GalT1 and its chaperone C1GalT1C1 also called Cosmc¹⁵.
51 Defective mutation in Cosmc is reported to affect the function of C1GalT1 for elongating Tn to
52 normal O-glycan structures^{15,16}. Furthermore, Tn is involved in IgA nephropathy (IgAN, also
53 known as Berger's disease) that is the most common glomerular disease in the world^{3,17,18}. A
54 large percentage of patients with IgAN progress to kidney failure, also called end-stage renal
55 disease (ESRD)^{3,17}. The cause of IgAN may involve the expression of Tn and STn on hinge
56 region of IgA1³.

57 Although Tn is structurally simple, identification of its glycosylation sites and the carrier
58 proteins in the complex samples is highly challenging due to the lack of suitable technology.
59 Limited information regarding Tn-glycosylation sites and carrier proteins hamper the
60 understanding of the role of Tn in cancer biology and the development of new strategies
61 targeting cancers. Current methods for mapping Tn-glycosylation sites include the use of VVA
62 lectin or hydrazide chemistry for the enrichment of Tn-glycopeptides, followed by LC-MS/MS
63 for site localization ^{19, 20}. Jurkat T cells expressing Tn and STn, due to the mutation in Cosmc,
64 are often used as a model system to evaluate the effectiveness of methods. Using VVA lectin
65 chromatography and ETD-MS2, Steentoft et al. identify 68 O-glycoproteins in Jurkat cells ¹⁹.
66 Zheng et al. use galactose oxidase to oxidize Tn followed by solid-phase capture using hydrazide
67 chemistry and release of Tn-glycopeptides using methoxyamine ²⁰. Subsequent analysis using
68 HCD-MS2 identifies 96 O-glycoproteins in three experiments with 87 glycosylation sites being
69 localized in the first experiment of Jurkat cells ²⁰. We, however, anticipate that about a thousand
70 Tn-glycosylation sites remain to be mapped in Jurkat cells because 1,295 O-linked glycosylation
71 sites are mapped in CEM cells, a human T cell line, using a method named EXoO developed in
72 previous study ²¹. It appears that the development of a technology capable of large-scale
73 mapping of Tn-glycosylation sites would be a significant advance in technology and cancer
74 biology.

75 Here, we introduce a new technology named EXoO-Tn that tags Tn and maps its
76 glycosylation sites in a large-scale. EXoO-Tn utilizes two highly specific enzymes in a one-pot
77 reaction for concurrent tagging of Tn and mapping of its glycosylation sites. The first enzyme is
78 glycosyltransferase C1GalT1, which catalyzes UDP-Gal to add a galactose to Tn. When
79 isotopically-labeled UDP-Gal(¹³C₆) is used, Gal(¹³C₆)-Tn is formed. The Gal(¹³C₆)-Tn has a

80 unique mass tag distinguishable to endogenous Gal-GalNAc and other glycans. The second
81 enzyme is an endoprotease named OpeRATOR, which cleaves at N-termini of Ser/Thr residues
82 occupied by the Gal(¹³C₆)-Tn to release site-containing Gal(¹³C₆)-Tn-glycopeptides with the
83 glycosylation sites positioning at the N-termini of peptide sequences. The two enzymes are
84 synergistically integrated with the use of solid-phase for optimal removal of contaminants and
85 efficient isolation of site-containing Gal(¹³C₆)-Tn-glycopeptides. A Proof-of-principle of EXoO-
86 Tn was developed using a synthetic Tn-glycopeptide. The performance of EXoO-Tn was
87 evaluated using Jurkat cells.

88

89 **Results**

90 **Principle of EXoO-Tn**

91 EXoO-Tn includes six steps (Fig. 1). (i) Digestion: proteins extracted from samples are digested
92 to peptides. Amino groups on the side chain of Lys residues are modified using guanidination on
93 C18 cartridge. (ii) Enrichment: Tn-glycopeptides are enriched using VVA lectin. (iii)
94 Conjugation: the enriched glycopeptides are conjugated to aldehyde-functionalized solid-phase
95 through amino groups at the peptide N-termini. (iv) Tn-engineering: Tn is catalyzed to Gal(¹³C₆)-
96 Tn using C1GalT1/C1GalT1C1 and UDP-Gal(¹³C₆). C1GalT1/C1GalT1C1 is specific to modify
97 Tn. The Gal(¹³C₆)-Tn has a unique mass that is distinguishable to endogenous Gal-GalNAc and
98 other glycans in the samples. (v) Release: site-containing Gal(¹³C₆)-Tn-glycopeptides are
99 specifically released from solid-phase using OpeRATOR enzyme, which cleaves N-termini of
100 Gal(¹³C₆)-Tn-occupied Ser/Thr residues. (vi) Analysis: the released glycopeptides are analyzed
101 using LC-MS/MS and software tools.

102 To show the feasibility of EXoO-Tn, a synthetic Tn-glycopeptide VPSTPPTPS(α -
103 GalNAc)PSTPPTPSPSC-NH₂ was used (Fig. 2A top left panel). The use of C1GalT1 and UDP-
104 Gal converted Tn to Gal-Tn produced a charge +2 Gal-Tn-glycopeptide at 1149.54 *m/z* (Fig. 2A
105 top middle panel), an increase of ~162 Da corresponding to the mass of a galactose compared to
106 its unmodified counterpart at 1068.51 *m/z* (Fig. 2A top left panel). The Gal-Tn-glycopeptide
107 could be digested by OpeRATOR to yield site-containing glycopeptide S(Gal-
108 Tn)PSTPPTPSPSC-NH₂ at 761.34 *m/z* and peptide VPSTPPTP at 795.42 *m/z* (Fig. 2A bottom
109 middle panel). To distinguish the newly engineered Gal-Tn from endogenous Gal-GalNAc and
110 other glycans, the UDP-Gal was substituted by an isotopically-labeled UDP-Gal(¹³C₆). The
111 Gal(¹³C₆) has all six carbon molecules in galactose labeled with carbon-13 featuring an
112 increment mass of 6 Da. The use of C1GalT1 and UDP-Gal(¹³C₆) successfully converted Tn to
113 Gal(¹³C₆)-Tn with a unique mass tag of 371 and yielded a charge +2 Gal(¹³C₆)-Tn-glycopeptide
114 at 1152.55 *m/z* (Fig. 2A top right panel), which had an increase of ~6 Da compared to its charge
115 +2 Gal-Tn counterpart at 1149.54 *m/z* (Fig. 2A top middle panel). The site-containing
116 glycopeptide S(Gal(¹³C₆)-Tn)PSTPPTPSPSC-NH₂ and peptide VPSTPPTP at 764.35 and 795.42
117 *m/z*, respectively, was generated after OpeRATOR digestion (Fig. 2A bottom right panel). The
118 Gal(¹³C₆)-Tn-glycopeptide had an increase of ~6 Da compared to its Gal-Tn or endogenous Gal-
119 GalNAc counterpart at 761.34 *m/z* (Fig. 2A bottom middle panel). Next, the MS/MS spectra of
120 site-containing Gal(¹³C₆)-Tn-glycopeptides were analyzed using HCD-MS2 to identify spectral
121 feature for improvement of confidence of identification. As an illustration, an MS/MS spectrum
122 of site-containing Gal(¹³C₆)-Tn-glycopeptide from analysis of Jurkat cells was shown (Fig. 2B).
123 A diagnostic oxonium ion generated by HCD fragmentation was observed at 372 *m/z* for the
124 Gal(¹³C₆)-Tn (Fig. 2B). The presence of the diagnostic oxonium ion at 372 *m/z* was utilized in

125 the data interpretation. The Gal(¹³C₆)-Tn-glycosylation site was informed to be the Thr residue at
126 the N-terminus of the identified peptide sequence (Fig. 2B). Other fragmentation ions in the
127 MS/MS spectrum, including oxonium ions, peptide b- and y-ions, and peptide ion supported the
128 identification of the glycopeptide (Fig. 2B). The analysis of glycopeptides demonstrated the key
129 enzymatic steps in EXoO-Tn to distinguish Tn from Gal-GalNAc and other glycans by isotopic
130 tagging using C1GalT1 and UDP-Gal(¹³C₆), and map Tn-glycosylation sites using OpeRATOR
131 and LC-MS/MS.

132

133 **Mapping site-specific Tn-glycoproteome in Jurkat cells**

134 Jurkat cells were analyzed to evaluate the performance of EXoO-Tn. With 1% FDR, 3,172
135 peptide-spectrum match (PSM) were assigned to 1,078 unique site-containing Gal(¹³C₆)-Tn-
136 glycopeptides that contained 1,011 unique peptide sequences (Fig. 3 and Supplementary Table
137 1). From the peptide sequence, we mapped 947 Gal(¹³C₆)-Tn-glycosylation sites from 480
138 glycoproteins (Fig. 3 and Supplementary Table 1). The diagnostic oxonium ion at 372 *m/z* was
139 detected in 96.4% of the assigned MS/MS spectra with an overall intensity being ten-fold lower
140 than that at 204 *m/z* (Fig. 4A and Supplementary Table 1). The detection of oxonium ion at 372
141 *m/z* in the assigned MS2 spectra supported the presence of Gal(¹³C₆)-Tn in the identified
142 glycopeptides (Supplementary Table 1). It was observed that, among the assigned PSMs,
143 approximately 89.2% glycopeptides were modified by a single Gal(¹³C₆)-Tn composition while
144 approximately 9.5 and 1.3% PSMs were modified by two or three Gal(¹³C₆)-Tn compositions,
145 respectively (Supplementary Table 1).

146

147 **Characterization of the site-specific Tn-glycoproteome in Jurkat cells**

148 Analysis of the glycosylation sites showed that Thr and Ser accounted for approximately 68.7%
149 and 31.3%, respectively. Motif analysis of ± 7 amino acids surrounding 946 glycosylation sites
150 found an overrepresentation of Pro residues at the +3 and -1 position (Fig. 4B). Two
151 glycosylation sites residing close to the protein N-termini were not used in the motif analysis.
152 Gene Ontology (GO) analysis of the identified glycoproteins found that integral component of
153 membrane, extracellular exosome, endoplasmic reticulum (ER), Golgi apparatus, cell surface,
154 and extracellular space were enriched for cellular component suggesting the presence of the
155 identified glycoproteins in the secretory pathway and on the cell surface (Fig. 4C). Next, the
156 relative position of the glycosylation sites in protein sequence was plotted and showed that
157 proteins MUC1 and versican core protein (VCAN) had the highest number of glycosylation sites
158 reaching 48 and 11, respectively (Fig. 4D middle panel). Besides, it was observed that the
159 frequency of the glycosylation site was relatively even across protein sequences with lower
160 frequency at protein termini (Fig. 4D top and bottom panels). Comparison of site-specific Tn-
161 glycoproteome identified by EXoO-Tn to two other methods^{19,20} (Supplementary Table 2 and 3)
162 revealed that 888 Tn-glycosylation sites from 398 glycoproteins were exclusively identified
163 using EXoO-Tn (Fig. 4E). Analysis of Jurkat cells established the effectiveness of EXoO-Tn to
164 map the site-specific Tn-glycoproteome in the complex sample.

165

166 **Discussion**

167 A new technology EXoO-Tn has been developed for large-scale mapping Tn-glycosylation sites
168 in the complex sample. EXoO-Tn has several advantages including (i) large-scale mapping of
169 Tn-glycosylation sites in the complex sample; (ii) a tagging strategy for distinguishing
170 engineered Tn from endogenous Gal-GalNAc and other glycans; (iii) concurrent tagging of Tn

171 and release of site-containing Tn-glycopeptides from solid-phase in a one-pot fashion; (iv)
172 applicable to analyze mucin-type O-linked glycoproteins; (v) no need of ETD for site
173 localization.

174 C1GalT1 is a natural enzyme with specificity for extending O-GalNAc to core 1 Gal-GalNAc
175 structure. OpeRATOR enzyme is utilized by bacteria to digest mucin glycoproteins in the gut
176 with a specificity at N-termini of Gal-GalNAc occupied Ser/Thr residues. The two enzymes
177 work synergistically to render EXoO-Tn the specificity for mapping Tn-glycosylation sites. It is
178 meritorious that Tn is tagged to have a unique mass and generate a diagnostic oxonium ion in the
179 MS2 spectrum. The unique mass tag and diagnostic oxonium ion are useful to improve the
180 confidence of identification. The use of solid-phase allows extensive washes that are essential to
181 remove other peptides and contaminants while enables further enrichment of site-containing
182 glycopeptides for LC-MS/MS analysis.

183 We mapped 947 Tn-glycosylation sites from almost 500 glycoproteins, a substantially large
184 number of site-specific Tn-glycoproteome, which demonstrated the effectiveness of EXoO-Tn
185 and supported that a large number of O-linked glycosylation sites could be mapped in cells.
186 Some site-containing Tn-glycopeptides may be too long or too short to be detected using EXoO-
187 Tn with trypsin digestion. Digestion of proteins using proteases with different specificities may
188 further increase the identification number of glycosylation sites in EXoO-Tn methodology. Also,
189 the identification of glycopeptides with two or three Gal(¹³C₆)-Tn compositions suggests many
190 more glycosylation sites in the peptide sequences supporting an even larger number of Tn-
191 glycosylation sites in Jurkat cells. Characterization of glycosylation sites and glycoproteins
192 identified in Jurkat cells revealed conserved features of protein O-linked glycosylation, including
193 consensus motif, cellular localization, and distribution of the relative position of glycosylation

194 sites across the protein sequences, a reminiscence of that seen in human kidney, serum, and T
195 cells in the previous study²¹. Given that Tn is prevalent in cancers and other diseases, EXoO-Tn
196 is anticipated to have broad translational and clinical utilities.

197

198 **Material and Methods**

199 **Tagging of Tn and mapping its glycosylation site using synthetic Tn-glycopeptide**

200 Synthetic Tn-glycopeptide VPSTPPTPS(α -GalNAc)PSTPPTPSPSC-NH₂ IgA1 hinge peptide
201 was purchased from Susses Research. In the workflow with sequential enzymatic treatments, five
202 μ g of glycopeptide in 50 mM Tris-HCl pH 7.4 was mixed with one μ g recombinant human
203 C1GalT1/C1GalT1C1 protein (R&D Systems, NM) in the presence of either 0.5 mM UDP-Gal
204 (Sigma-Aldrich) or 0.5 mM UDP-Gal¹³C₆ (Omicron Biochemicals, Inc., IN) at 37°C for 16
205 hours. After incubation, half of each sample was subjected to digestion using five units of
206 OpeRATOR (Genovis Inc, Cambridge, MA) at 37°C for 16 hours. The glycopeptides were
207 desalted using C18 ZipTip (Millipore Sigma), dried using speed-vac, and resuspended in 0.1%
208 TFA. In the concurrent one-pot enzymatic treatment that was used in all experiments described
209 below, enzymes including C1GalT1/C1GalT1C1, OpeRATOR, and substrate i.e. UDP-Gal or
210 UDP-Gal¹³C₆ were added at the same time using the amount as described in the above sequential
211 enzymatic workflow and incubated at 37°C for 16 hours before C18 desalting and LC-MS/MS
212 analysis.

213

214 **Extraction of site-containing Tn-glycopeptides from Jurkat cells**

215 Jurkat Clone E6-1 (NIH AIDS Reagent Program) were cultured and expanded in RPMI 1640
216 supplemented with 10% fetal bovine serum (FBS), 100 units of penicillin, and 100 μ g of

217 streptomycin. The cells were collected, washed three times in the ice-cold PBS and lysed in 8 M
218 urea/500 mM ammonia bicarbonate. The cell lyse was sonicated and centrifuged at 16,000 g to
219 remove particles. Protein concentration was determined using a protein BCA assay. Twenty
220 milligrams of proteins were reduced in 5 mM DTT at 37°C for 1 hour and alkylated in 10 mM
221 iodoacetamide at room temperature (RT) for 40 min in the dark. The samples were then diluted
222 five-fold using 100 mM ammonia bicarbonate buffer. Trypsin was added to the samples with an
223 enzyme/protein ratio of 1/40 w/w. After incubation at 37°C for 16 hours, lysine residues were
224 guanidination-modified, and peptides were desalted using C18 cartridges (Waters, Milford, MA),
225 as described in the previous study²¹. The peptides were dried using speed-vac, resuspended in
226 PBS with α 2-3,6,8 neuraminidase (New England Biolabs, Ipswich, MA), and incubated at 37°C
227 for 16 hours. Four-hundred microliters agarose bound Vicia Villosa Lectin (VVA) (50% slurry,
228 Vector Laboratories, Burlingame, CA) were washed twice using water, added to peptides and
229 incubated at RT for 16 hours with rotation. The VVA agarose was gently washed with 1X PBS
230 for three times. Bound glycopeptides were eluted using 4 M urea/100 mM Tris-HCl pH
231 7.4/400mM GalNAc (Sigma-Aldrich) at RT for 30 min with shaking. The eluted glycopeptides
232 were desalted using C18 cartridge and conjugated to AminoLink resin (Pierce, Rockford, IL) as
233 described previously²¹. Briefly, the pH of C18 elute containing glycopeptides was neutralized to
234 approximately pH 7 using two volume of 10X PBS. The solution was mixed with resin (100 μ g
235 peptide/100 μ l resin, 50% slurry) and 50 mM sodium cyanoborohydride (NaCNBH₃) at RT for a
236 minimal of 4 hours or overnight with rotation. Unreacted groups on resin were blocked using 1M
237 Tris-HCl buffer (pH7.4) with 50 mM NaCNBH₃ at RT for 30 min with rotation. The resin was
238 sequentially washed using 50% ACN, 1.5 M NaCl, and 50 mM Tris-HCl buffer (pH 7.4). To tag
239 and release Tn-glycopeptides, a solution (50 μ l) containing 10 μ g of C1GalT1/C1GalT1C1, 0.5

240 mM UDP-Gal¹³C₆, and 2000 units of OPERATOR was added to the resin and incubated at 37°C
241 for 16 hours. The released glycopeptides in the solution were collected twice using 400 µl of 50
242 mM Tris-HCl buffer (pH 7.4). Glycopeptides in the collected solution were combined, desalted
243 using C18 cartridge, dried using speed-vac, and resuspended in 0.1% TFA. The peptides were
244 fractionated using HPLC and concatenated to eight fractions before LC-MS/MS analysis.

245

246 **LC-MS/MS analysis**

247 One microgram of glycopeptides was analyzed on a Fusion Lumos mass spectrometer with an
248 EASY-nLC 1200 system or an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher
249 Scientific, Bremen, Germany). The mobile phase flow rate was 0.2 µl/min with 0.1% FA/3%
250 acetonitrile in water (A) and 0.1% FA/90% acetonitrile (B). The gradient profile was set as
251 follows: 6% B for 1 min, 6–30% B for 84 min, 30–60% B for 9 min, 60–90% B for 1 min, 90%
252 B for 5 min and equilibrated in 50% B, flow rate was 0.5 µL/min for 10 min. MS analysis was
253 performed using a spray voltage of 1.8 kV. Spectra (AGC target 4×10^5 and maximum injection
254 time 50 ms) were collected from 350 to 1800 m/z at a resolution of 60 K followed by data-
255 dependent HCD MS/MS (at a resolution of 50 K, collision energy 36, AGC target of 2×10^5 and
256 maximum IT 250 ms) of the 15 most abundant ions using an isolation window of 0.7 m/z.
257 Include charge state was 2-6. The fixed first mass was 110 m/z. Dynamic exclusion duration was
258 45 s.

259

260 **Database search of site-containing Tn-glycopeptides**

261 A UniProt human protein database (71,326 entries, downloaded October 19, 2017) was used to
262 generate a peptide database with 26,067,074 non-redundant peptide entries using the method as

263 described in the previous study ²¹. Briefly, a randomized decoy database using The Trans-
264 Proteomic Pipeline (TPP) ²² was generated and concatenated with the target database. The
265 concatenated database was digested with trypsin and then OpeRATOR in silico. Peptides with
266 Ser or Thr residues and lengths from 6 to 46 amino acids were used. SEQUEST in Proteome
267 Discoverer 2.2 (Thermo Fisher Scientific) was used to search with variable modification:
268 oxidation (M), Gal¹³C₆(1)HexNAc(1) (S/T), Hex(1)HexNAc(1) (S/T) and HexNAc (S/T) and
269 static modification: carbamidomethylation (C) and guanidination (K). FDR was set at 1% using
270 Percolator. Only MS/MS scans with oxonium ion at 204, and two of the other oxonium ions were
271 kept. Assignments with XCorr score below one were removed. MS/MS spectra were manually
272 studied and inspected using spectral viewer in Proteome Discoverer to identify the spectral
273 feature and ensure the confidence of identification.

274

275 **Bioinformatics**

276 Software pLogo was used to reveal motif for Tn-glycosylation sites ²³ surrounding by 15 amino
277 acids in length with the central amino acids being the sites. The Database for Annotation,
278 Visualization and Integrated Discovery (DAVID) and UniProt (<http://www.uniprot.org>) were
279 used for Gene Ontology (GO) analysis ²⁴. Python (version 2.7) is used to analyze the data and
280 generate the figures, including the relative position of Tn-glycosylation sites in protein sequence,
281 radar charts, unsupervised hierarchical clustering, and box plot.

282

283 **Data Availability**

284 The LC-MS/MS data have been deposited to the PRIDE partner repository ²⁵ with the dataset
285 identifier: project accession: PXD014390

286 Reviewer account details:

287 Username: reviewer03140@ebi.ac.uk

288 Password: tZVBNHhu

289

290 **Acknowledgment**

291 We acknowledge Dr. Kyung-Cho Cho for maintenance of Mass Spectrometer. This work was
292 supported by the National Cancer Institute, the Early Detection Research Network (EDRN,
293 U01CA152813), the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA210985),
294 the National Institute of Allergy and Infectious Diseases (R21AI122382), and by amfAR, The
295 Foundation for AIDS Research on Bringing Bioengineers to Cure HIV (Grant amfAR 109551-
296 61-RGRL). The following reagent was obtained through the NIH AIDS Reagent Program,
297 Division of AIDS, NIAID, NIH: Jurkat Clone E6-1 from Dr. Arthur Weiss (cat# 177)²⁶.

298

299 **Affiliations**

300 Department of Pathology, Johns Hopkins University, Baltimore, Maryland, USA.

301 Weiming Yang, Minghui Ao, Angellina Song, Yuanwei Xu, and Hui Zhang

302

303 **Contributions**

304 W.Y. and H.Z. conceived the concept and wrote the manuscript; W.Y., A.S., and Y.X. conducted
305 experiments and data analysis; M.A performed programming, advanced data analysis, and
306 bioinformatics.

307

308 **Competing financial interests**

309 The authors declare no competing financial interests.

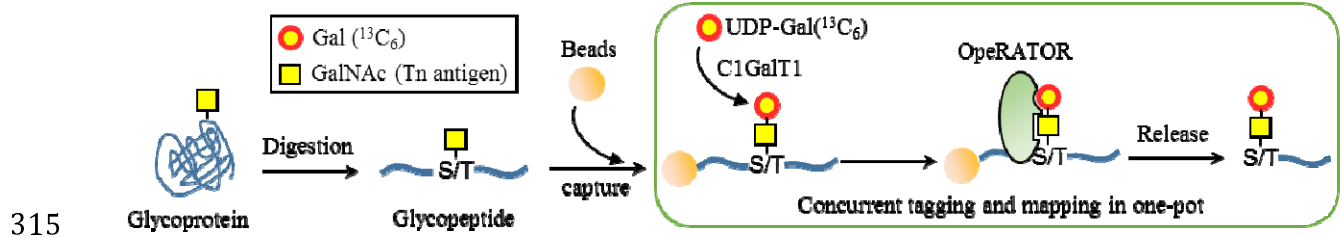
310

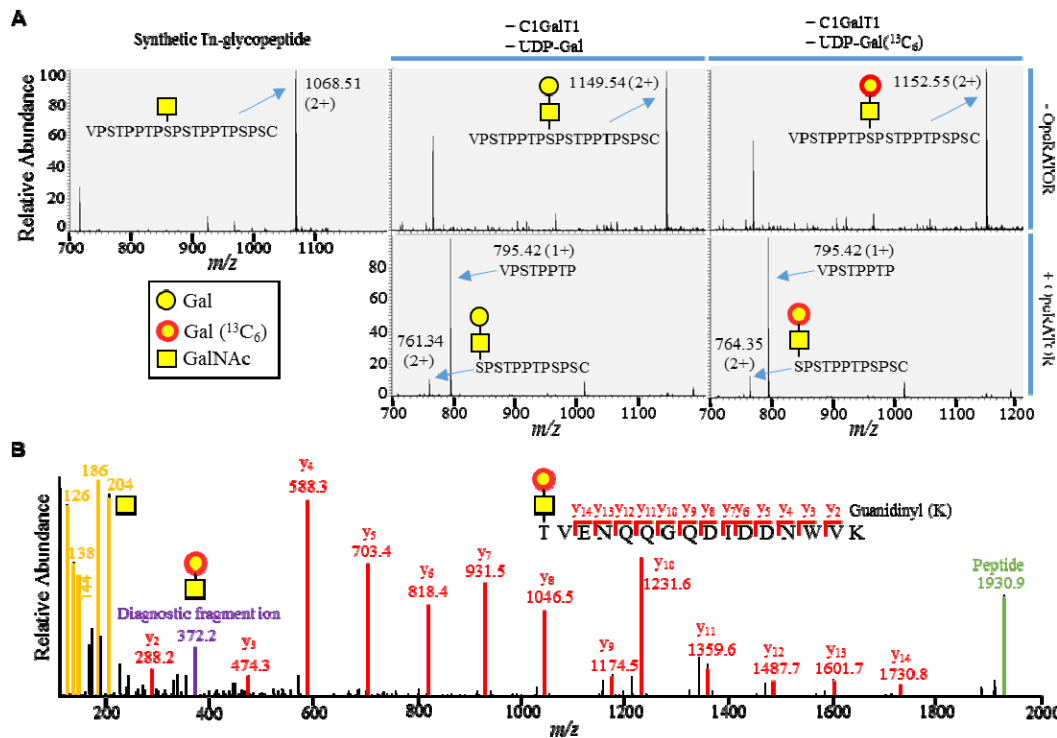
311 **Corresponding author**

312 Correspondence to: Weiming Yang

313

314 **Figures and figure legend**

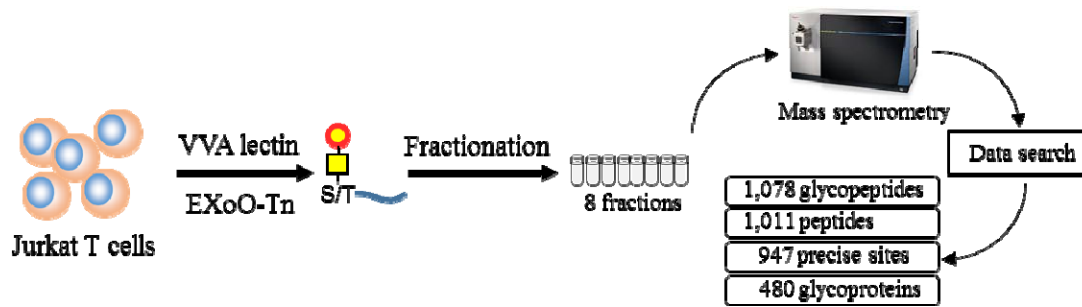




317

318 Figure 2 Mapping Tn-glycosylation sites by integrating Tn-engineering and OpeRATOR
 319 digestion.

320 A. OpeRATOR digestion of Gal- and Gal(¹³C₆)-Tn-glycopeptide after Tn was tagged using
 321 C1GalT1 with UDP-Gal or UDP-Gal(¹³C₆). Top left panel: the synthetic Tn-glycopeptide before
 322 treatments. Top middle panel: conversion of Tn to Gal-Tn using C1GalT1 and UDP-Gal. Bottom
 323 middle panel: OpeRATOR digestion of the Gal-Tn-glycopeptide generated in the top middle
 324 panel produced site-containing glycopeptide S(Gal-Tn)PSTPPTPSPSC-NH₂ and peptide
 325 VPSTPPTP. Top right panel: conversion of Tn to Gal(¹³C₆)-Tn using C1GalT1 and UDP-
 326 Gal(¹³C₆). Bottom right panel: OpeRATOR digestion of the Gal(¹³C₆)-Tn-glycopeptide
 327 engineered in the top right panel yielded site-containing glycopeptide S(Gal(¹³C₆)-
 328 Tn)PSTPPTPSPSC-NH₂ and peptide VPSTPPTP. B. HCD-MS2 spectrum of site-containing
 329 Gal(¹³C₆)-Tn-glycopeptide identified in Jurkat cells. A diagnostic oxonium ion at 372 m/z
 330 corresponding to fragmentation ion of Gal(¹³C₆)-Tn was colored in purple.



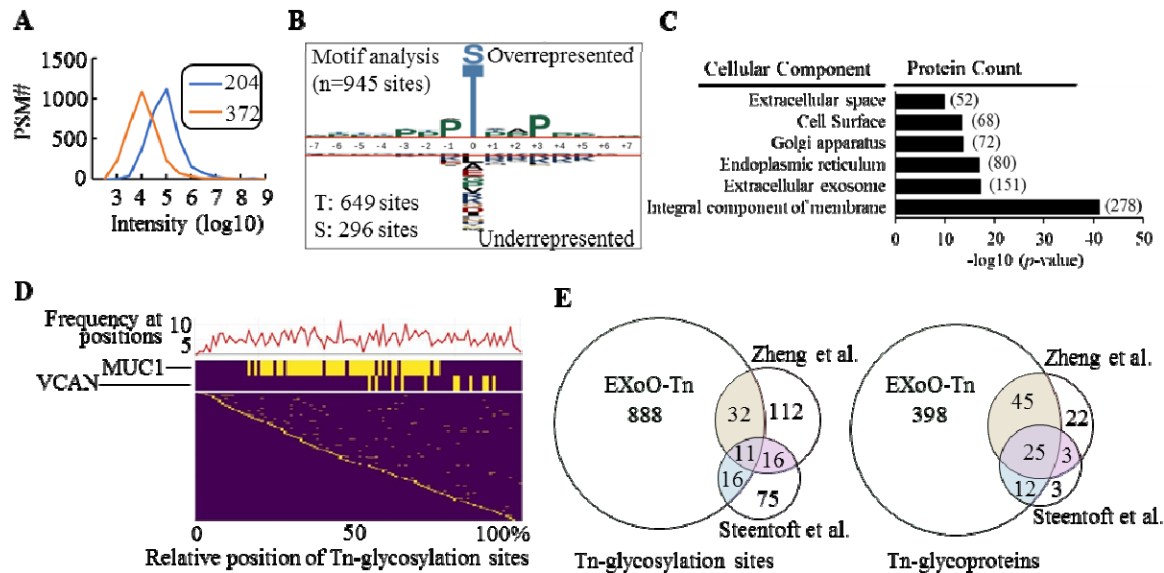
331

332 Figure 3 A Schematic workflow for identification of site-specific Tn-glycoproteome in Jurkat

333 cells.

334

335



336

337 Figure 4 Characteristics of site-specific Tn-glycoproteome in Jurkat cells.

338 A. The overall intensity of oxonium ions at 204 m/z in the assigned PSMs. The overall
 339 intensity of oxonium ion at 372 m/z was 10-fold less than that of 204 m/z . B. Motif analysis
 340 revealed the conserved motif of Tn-glycosylation sites. C. GO analysis revealed cellular
 341 components for Tn-glycoproteome. D. Analysis of the relative position of Tn-glycosylation sites
 342 in protein sequences revealed that the frequency of Tn-glycosylation distributed evenly across
 343 protein sequences with lower frequency at protein termini. E. Comparison of O-linked
 344 glycosylation sites and glycoproteins identified in this and other studies^{19,20}.

345

346 References

- 347 1. Julien, S., Videira, P.A. & Delannoy, P. Sialyl-Tn in cancer: (how) did we miss the
348 target? *Biomolecules* **2**, 435-466 (2012).
- 349 2. Munkley, J. The Role of Sialyl-Tn in Cancer. *International journal of molecular*
350 *sciences* **17**, 275 (2016).
- 351 3. Ju, T. et al. Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease
352 markers. *Proteomics. Clinical applications* **7**, 618-631 (2013).
- 353 4. Kudelka, M.R., Ju, T., Heimbürg-Molinario, J. & Cummings, R.D. Simple sugars to
354 complex disease--mucin-type O-glycans in cancer. *Advances in cancer research* **126**,
355 53-135 (2015).
- 356 5. Slovin, S.F. et al. Fully synthetic carbohydrate-based vaccines in biochemically
357 relapsed prostate cancer: clinical trial results with alpha-N-acetylgalactosamine-O-
358 serine/threonine conjugate vaccine. *Journal of clinical oncology : official journal of*
359 *the American Society of Clinical Oncology* **21**, 4292-4298 (2003).
- 360 6. Itzkowitz, S.H., Bloom, E.J., Lau, T.S. & Kim, Y.S. Mucin associated Tn and sialosyl-Tn
361 antigen expression in colorectal polyps. *Gut* **33**, 518-523 (1992).
- 362 7. Inoue, M., Ton, S.M., Ogawa, H. & Tanizawa, O. Expression of Tn and sialyl-Tn
363 antigens in tumor tissues of the ovary. *American journal of clinical pathology* **96**,
364 711-716 (1991).
- 365 8. Wei, H. et al. Glycoprotein screening in colorectal cancer based on differentially
366 expressed Tn antigen. *Oncology reports* **36**, 1313-1324 (2016).
- 367 9. Nakagoe, T. et al. Prognostic value of circulating sialyl Tn antigen in colorectal
368 cancer patients. *Anticancer research* **20**, 3863-3869 (2000).
- 369 10. Tsuchiya, A. et al. Prognostic Relevance of Tn Expression in Breast Cancer. *Breast*
370 *cancer* **6**, 175-180 (1999).
- 371 11. Ohno, S. et al. Expression of Tn and sialyl-Tn antigens in endometrial cancer: its
372 relationship with tumor-produced cyclooxygenase-2, tumor-infiltrated lymphocytes
373 and patient prognosis. *Anticancer research* **26**, 4047-4053 (2006).
- 374 12. Posey, A.D., Jr. et al. Engineered CAR T Cells Targeting the Cancer-Associated Tn-
375 Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* **44**,
376 1444-1454 (2016).
- 377 13. Wilkie, S. et al. Retargeting of human T cells to tumor-associated MUC1: the
378 evolution of a chimeric antigen receptor. *Journal of immunology* **180**, 4901-4909
379 (2008).
- 380 14. Maher, J. et al. Targeting of Tumor-Associated Glycoforms of MUC1 with CAR T Cells.
381 *Immunity* **45**, 945-946 (2016).
- 382 15. Ju, T. et al. Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc.
383 *Cancer research* **68**, 1636-1646 (2008).
- 384 16. Hofmann, B.T. et al. COSMC knockdown mediated aberrant O-glycosylation
385 promotes oncogenic properties in pancreatic cancer. *Molecular cancer* **14**, 109
386 (2015).
- 387 17. Moran, S. & Cattran, D.C. IgA nephropathy: un update. *Minerva medica* (2019).
- 388 18. Berger, J. & Hinglais, N. [Intercapillary deposits of IgA-IgG]. *Journal d'urologie et de*
389 *nephrologie* **74**, 694-695 (1968).

- 390 19. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-
391 glycoengineered SimpleCell lines. *Nature methods* **8**, 977-982 (2011).
- 392 20. Zheng, J., Xiao, H. & Wu, R. Specific Identification of Glycoproteins Bearing the Tn
393 Antigen in Human Cells. *Angewandte Chemie* **56**, 7107-7111 (2017).
- 394 21. Yang, W., Ao, M., Hu, Y., Li, Q.K. & Zhang, H. Mapping the O-glycoproteome using site-
395 specific extraction of O-linked glycopeptides (EXoO). *Mol Syst Biol* **14**, e8486 (2018).
- 396 22. Deutsch, E.W. et al. Trans-Proteomic Pipeline, a standardized data processing
397 pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clinical*
398 *applications* **9**, 745-754 (2015).
- 399 23. O'Shea, J.P. et al. pLogo: a probabilistic approach to visualizing sequence motifs.
400 *Nature methods* **10**, 1211-1212 (2013).
- 401 24. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of
402 large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
- 403 25. Vizcaino, J.A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic*
404 *acids research* **44**, D447-456 (2016).
- 405 26. Weiss, A., Wiskocil, R.L. & Stobo, J.D. The role of T3 surface molecules in the
406 activation of human T cells: a two-stimulus requirement for IL 2 production reflects
407 events occurring at a pre-translational level. *Journal of immunology* **133**, 123-128
408 (1984).
409