Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics

Authors
Yoshiki Narimatsu, Hiren J. Joshi, Katrine T. Schjoldager, John Hintze, Adnan Halim, Catharina Steentoft, Rebecca Nason, Ulla Mandel, Eric P. Bennett, Henrik Clausen, and Sergey Y. Vakhrushev

Correspondence
yoshiki@sund.ku.dk; seva@sund.ku.dk

In Brief
The hereto most comprehensive characterization of nonredundant contributions of individual GalNAc-T isoforms to the O-glycoproteome of the human HEK293 cell using quantitative differential O-glycoproteomics on a panel of isogenic HEK293 cells with knockout of GalNAc-T genes (GALNT1, T2, T3, T7, T10, or T11). The engineered HEK293 cell panels are useful for expression and analysis of the regulation of O-glycosylation and enables wider studies of the biological functions of GalNAc-T isoform-specific glycosylation.

Highlights
- A panel of HEK293 isogenic cell lines with knockout of GALNT genes.
- Identification of nonredundant O-glycosylation sites regulated by specific GalNAc-T isoforms.
- GalNAc-T7 and T10 contribute to follow-up activity in regions of high density O-glycosylation.
- GalNAc-T11 specifically controls O-glycosylation of specific linker regions in the low-density lipoprotein receptor related proteins.
Most proteins trafficking the secretory pathway of metazoan cells will acquire GalNAc-type O-glycosylation. GalNAc-type O-glycosylation is differentially regulated in cells by the expression of a repertoire of up to twenty genes encoding polypeptide GalNAc-transferase isoforms (GalNAc-Ts) that initiate O-glycosylation. These GalNAc-Ts orchestrate the positions and patterns of O-glycans on proteins in coordinated, but poorly understood ways - guided partly by the kinetic properties and substrate specificities of their catalytic domains, as well as by modulatory effects of their unique GalNAc-binding lectin domains. Here, we provide the hereto most comprehensive characterization of nonredundant contributions of individual GalNAc-T isoforms to the O-glycoproteome of the human HEK293 cell using quantitative differential O-glycoproteomics on a panel of isogenic HEK293 cells with knockout of GalNAc-T genes (GALNT1, T2, T3, T7, T10, or T11). We confirm that a major part of the O-glycoproteome is covered by redundancy, whereas distinct O-glycosite subsets are covered by nonredundant GalNAc-T isofrm-specific functions. We demonstrate that the GalNAc-T7 and T10 isoforms function in follow-up of high-density O-glycosylated regions, and that GalNAc-T11 has highly restricted functions and essentially only serves the low-density lipoprotein-related receptors in linker regions (C^3XXCT^3) between the ligand-binding repeats. Molecular & Cellular Proteomics 18: 1396–1409, 2019. DOI: 10.1074/mcp.RA118.001121.

O-glycosylation of the N-acetylgalactosamine (GalNAc)-type is one of the most abundant and diverse forms of protein glycosylation, and it is uniquely positioned in the secretory pathway to fine-tune protein function (1, 2). GalNAc-type O-glycosylation (hereafter simply O-glycosylation) is controlled by many polypeptide GalNAc-transferase (GalNAc-T) isoenzymes encoded for by up to 20 distinct GALNT genes in mammals that catalyze the addition of GalNAc residues to select serine and threonine (and possibly tyrosine) residues. GalNAc-T isoenzymes have distinct, albeit partly overlapping, peptide substrate specificities and kinetic properties, they are differentially expressed in cells and tissues, and their expression patterns change during cellular maturation, differentiation, and malignant transformation (1, 3, 4). The repertoire of isoenzymes expressed in cells direct the positions and patterns of O-glycans found on proteins (5), but our insight into the specific contributions of individual isoenzymes and the seemingly coordinated process leading to proficient O-glycosylation of both isolated and clustered glycosites is still highly limited. GalNAc-Ts are unique among metazoan glycosyltransferases in that they employ a C-terminal GalNAc-binding lectin domain to modulate the substrate specificity and properties of the catalytic domain, which is predicted to enable follow-up glycosylation in high-density O-glycosylation regions (6, 7). The lectin domains coordinate glycosylation of distant glycosites (± 8–10 residues of initial O-glycosites) with isofrm-specific orientation (7, 8), and recently the molecular basis for this has been shown to involve combined docking of the partially glycosylated substrate into the lectin and catalytic domain (9). Moreover, a subset of the GalNAc-T isoenzymes selectively or exclusively recognizes acceptor sites with GalNAc residues found immediately adjacent (10), and structural studies demonstrate that the catalytic domain accommodates the first introduced GalNAc residue (10, 11). Most of our current understanding is, however, still based largely on in vitro analyses with synthetic peptide and glycopeptide substrates (12, 13).

Recently, we developed a quantitative O-glycoproteomics strategy for sensitive mapping of contribution of individual GalNAc-Ts with identification of their nonredundant functions (5). This strategy is partly based on genetic simplification of the O-glycan structures produced in cells with the so-called “SimpleCell” strategy, where the second step in the biosynthesis of O-glycans is eliminated by knockout (KO) of the
private chaperone Cosmc for the core1 synthase C1GalT1 (14). This facilitates the lectin enrichment required for sensitive mass spectrometric identification and quantification of glycopeptides (15, 16). The strategy is further based on isogenic SimpleCells with KO of individual GALNT genes and comparative analysis of O-glycoproteomes using stable isotope dimethyl labeling of tryptic digests, lectin enrichment and sensitive mass spectrometry (5, 17, 18). In a previous study, we applied the strategy to human liver HepG2 cells targeting the two most abundantly expressed GALNT1 and T2 genes, and demonstrated that these two isoforms each serve a small unique subset of substrates, whereas the majority of the O-glycosylation capacity was predicted to be redundantly shared (5). Similarly, we found that the two close isoforms, GaINAc-T3 and T6, predicted to have highly similar functions selectively served a minor set of unique substrates (19, 20). These results are in good agreement with the past extensive in vitro studies performed (12, 13), indicating that the strategy can be used to explore the contribution of GaINAc-T isoenzymes to the O-glycoproteome.

Here, we took the strategy a step further, and used precision gene engineering to develop a comprehensive panel of isogenic HEK293 SimpleCell (SC) and corresponding wildtype (WT) cell lines with different repertoires of the six major GALNT genes that are expressed, to explore regulation of the O-glycoproteome in detail in a human cell widely used for recombinant expression of glycoproteins. We confirm that a major part of the O-glycoproteome is covered by redundant functions of GaINAc-T isoenzymes, and we identify distinct O-glycosites that are regulated by specific GaINAc-T isoforms. We present the first evidence in cells that GaINAc-T7 and T10 serve to follow-up initial partial O-glycosylation in regions with clustered O-glycosites. Moreover, we find that the GaINAc-T11 isoform exhibits the most restricted nonredundant functions found so far for GaINAc-T isoforms, and T11 essentially only controls O-glycosylation of the low-density lipoprotein receptor (LDLR) and related proteins (LRPCs) specifically in the C\(^6\)XXXTC\(^1\) sequence motif in the linker regions of LDLR class A repeats. The panel of HEK293 WT isogenic cell lines with knockout of GALNTs presented here are useful for dissection and validation of biological functions of site-specific O-glycosylation.

**EXPERIMENTAL PROCEDURES**

**Generation of GALNT Gene Knockout Cell Lines**—Gene targeting using Zinc finger nucleases (ZFNs) were performed in the HEK293\(^{SC}\) cell line with COSMC KO established previously (14) and in HEK293 WT. Cells were maintained in Dulbecco’s Modified Eagle Medium supplemented with 10% FBS, 1% glutamax. ZFN GFP/Crimson fusion constructs were transfected by electroporation using Amaxa Nucleofector 2B system (Lonza, Switzerland), and GFP/Crimson double positive cells were enriched by FACS sorting. After 1–2 weeks of culture bulk sorted cell populations were further single cell sorted for GFP/Crimson negative population. KO clones with frameshift mutations were identified by Indel Detection Amplification Assay (IDAA), as previously described (21). All ZFN target sites and the primers used for IDAA are listed in supplemental Table S1. All selected clones were further verified by TOPO cloning and Sanger sequencing for characterization of mutations introduced.

**Immunocytochemistry**—HEK293 cells were grown on sterile cover slides and fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100, and blocked with 3% BSA. Fixed cells were incubated with monoclonal antibodies (MAbs) to GaINAc-T1 (4D8), T2 (4C4), T3 (2D10), T7 (8B8), or T11 (1B2) (22, 23) overnight at 4 °C, followed by FITC-conjugated rabbit anti-mouse Ig (Dako) for 45 min, and mounted with ProLong Gold antifade reagent (DAPI) (Invitrogen, Carlsbad, CA). Fluorescence microscopy was performed using a Zeiss Axioskop 2 plus with an AxioCam MR3 (Axioskop, Zeiss, Oberkochen, Germany).

**Dimethyl Labeling and Lectin Weak Affinity Chromatography**—The sample preparation for O-glycoproteomics was previously described (9). Briefly, packed cell pellets (0.5 ml) from HEK293\(^{SC}\) = GALNT KO were lysed with 0.1% RapiGest (Waters Corporation, Milford, MA) or 200 ml cell culture supernatant were applied to a 1 ml Viscosa Villosa lectin (VVA) lectin agarose (Vector Laboratories, Inc., Burlingame, CA) column and bound glycoproteins were eluted by heating a slurry of the agarose in 0.02% RapiGest. Cleared lysates and secretome samples were heated for 10 min at 80 °C, followed by reduction (5 mM DTT, 60 °C, 30 min), alkylation (10 mM iodoacetamide, RT, 30 min), and digestion with trypsin (25 μg/sample) (Roche, Hvidovre, Denmark) (37 °C, ON). Cleared acidified digestes were loaded onto equilibrated SepPak C18 cartridges (Waters) followed by 0.1% TFA wash and labeled on-column by adding 30 mM NaBH\(_3\)CN and 0.2% formic acid (light-labeling, L) or 30 mM NaBH\(_4\)CN and 0.2% d-formaldehyde (medium-labeling, M). Columns were washed using 0.1% FA and eluted with 50% MeOH in 0.1% FA. Labeled GaINAc-glycopeptides were separated from nonglycosylated peptides using VVA-agarose LWAC as previously described (5).

**Quantitative Differential O-glycoproteomic Strategy**—The glycopeptide quantification based on M/L isotope labeled doublet ratios was evaluated to estimate a meaningful cut-off ratio for substantial changes (5). The labeled glycopeptides produced doublets with varying ratios of the isotopic ions as well as a significant number of single precursor ions without evidence of ion pairs. Labeled samples from HEK293\(^{SC}\) and HEK293\(^{KO}\) with KO of individual GALNT genes were mixed 1:1 and subjected to LWAC separation. The distribution of labeled peptides from the LWAC flow-through showed that the quantitated peptide M/L ratios were normally distributed with 99% falling within ±1 (Log\(_{10}\)). We selected doublet with less/more than −1/1 (Log\(_{10}\)) value and singlets as candidates for isomform-specific O-glycosylation events.

**Isoelectric Focusing**—LWAC fractions most enriched in glycopeptides were pooled together, dried by vacuum centrifugation, reconstituted in IPG rehydration buffer, and submitted to IEF fractionation as previously described (18). Isoelectric focusing (IEF) was performed by a 3100 OFFGEL fractionator (Agilent Technologies, La Jolla, CA) using pH 3–10 strips (GE Healthcare, Hertfordshire, UK) 12 fractions were collected and desalted by custom Stage Tips (C18 sorbent from Empore 3 ml) and submitted to LC-MS/MS analysis.

**Mass Spectrometry**—EASY-nLC 1000 UHPLC (Thermo Scientific, Odense, DK) interfaced via nanoSpray Flex ion source to an LTQ-Orbitrap Velos Pro spectrometer (Thermo Scientific, San Jose, CA) was used for the glycoproteomic study. The nLC was operated in a single analytical column set up with PicoFrit Emitters (New Objectives, 75 μm inner diameter) packed with Reprosil-Pure-AQ C18...
Deconstruction of O-glycosylation in HEK293 Cells

phase (Dr. Maisch, 1.9-μm particle size, 19–21 cm column length). Each sample was injected onto the column and eluted with either a 2 h gradient (2–20% B in 95 min, 20–80% B in 10 min, and 80% B for 15 min) or a 4 h gradient (2–20% B in 215 min, 20–80% B in 10 min, and 80% B for 15 min) at 200 nl/min (Solvent A: 100% H2O; Solvent B: 100% acetonitrile; both containing 0.1% (v/v) formic acid).

A precursor MS1 scan (m/z 350–1,700) of intact peptides was acquired in the Orbitrap at a nominal resolution setting of 30,000. The five most abundant multiply charged precursor ions in the MS1 spectrum at a minimum MS1 signal threshold of 50,000 was triggered for sequential Orbitrap HCD-MS2 and ETD-MS2 (m/z of 100–2,000). MS2 spectra were acquired at a resolution of 7,500 for HCD MS2 and 15,000 for ETD MS2. Activation times were 30 and 200 ms for HCD and ETD fragmentation, respectively; isolation width was 4 mass units, and 1 microscan was collected for each spectrum. Automatic gain control targets were 1,000,000 ions for Orbitrap MS1 and 100,000 for MS2 scans, and the automatic gain control for fluorescent ion used for ETD was 300,000. Supplemental activation (20%) of the charge-reduced species was used in the ETD analysis to improve fragmentation. Dynamic exclusion for 60 s was used to prevent repeated analysis of the same components. Polysiloxane ions at m/z 445,12003 were used as a lock mass in all runs. The mass spectrometry glycoproteomics raw data and annotated spectra have been deposited to the ProteomeXchange Consortium (http://proteomexchange.org) (24) via the PRIDE partner repository with the data set identifier PXD009955.

Experimental Design and Statistical Rationale — The total number of samples (WT & KO pair) analyzed and described was 14. Each sample was represented by 12 IEF fractions. Each fraction was analyzed one time and no replicate analyses were performed because this is an exploratory study to identify targets for further validation. Median quantification ratios were calculated for each identified glycopeptide species.

Data Analysis — Data processing was performed using Proteome Discoverer 1.4 software (Thermo Scientific) using Sequest HT Node as previously described (5).

Briefly, all spectra were initially searched with trypsin full cleavage specificity, filtered according to the confidence level (medium, low and unassigned) and further searched with the semi-specific enzymatic cleavage. The maximum number of missed cleavage sites was set to 2. In all cases the precursor mass tolerance was set to 6 ppm and fragment ion mass tolerance to 20 mnu. Carbamidomethylation on cysteine residues was used as a fixed modification. Methionine oxidation and HexNAc attachment to serine, threonine and tyrosine were used as variable modifications for ETD MS2. All HCD MS2 were preprocessed as described (25) and searched under the same conditions mentioned above using only methionine oxidation as variable modification. All spectra were searched against a concatenated forward/reverse human-specific database (UniProt, January 2013, containing 20,232 canonical entries and another 251 common contaminants) using a target false discovery rate (FDR) of 1%. FDR was calculated using target decoy PSM validator node. The resulting list was filtered to include only peptides with glycosylation as a modification. Glycopeptide M/L ratios were determined using dimethyl 2plex method as previously described (5).

RESULTS

Generation of a Panel of Isogenic HEK293SC with Different Repertoires of GalNAc-Ts — The cellular repertoire of GalNAc-T isoforms determines the O-glycosylation capacity and the part of the cellular proteome that undergoes O-glycosylation (5). The repertoire of expressed GALNTs in HEK293 cells was analyzed by RNAseq as well as by immunochemistry (ICC) for isoforms for which validated monoclonal antibodies (mAbs) were available (Fig. 1A and supplemental Fig. S1). RNAseq analysis indicated that GALNT1, T2, T3, T7, T10, and T11 were the predominant isoforms expressed, whereas GALNT4, T6, T12, T13, T16, and T18 were barely detectable. Expression of the former isoforms were confirmed by ICC with specific mAbs, and for the latter group no detectable expression was found by ICC, although GalNAc-T13 and T18 were not analyzed because mAbs were not available (supplemental Fig. S1).

To probe the contributions of the major GalNAc-Ts to the O-glycoproteome we generated a panel of isogenic HEK293SC with KO of the six highly expressed GALNTs including HEK293SCAT1, HEK293SCAT2, HEK293SCAT3, HEK293SCAT7, HEK293SCAT10, and HEK293SCAT11 (Fig. 1B). We used ZFN gene targeting and generated two independent clones of each and characterized the indels introduced by IDAA and Sanger sequencing (supplemental Table S1). We also generated the double GALNT7/T10 KO HEK293SCAT7T10, because T7 and T10 are predicted to present closely related isoenzymes with similar functions (26–28). The HEK293 GALNT KO clones were further validated by immunocytoology for loss of the targeted GALNT, and immunostaining with available mAbs to ten human GalNAc-Ts revealed no apparent changes in expression of other GALNT genes (supplemental Fig. S2). Finally, all GALNT KOs were also developed in HEK293VT to provide availability of HEK293 cells with different GALNT repertoires and capacities for initiation of O-glycosylation, but with normal capacity for elongation and capping of O-glycans. The complete collection of isogenic cell lines generated for this study is summarized in supplemental Table S1.

Quantitative Differential O-glycoproteome Analysis — The differential O-glycoproteome strategy using stable isotope dimethyl labeling of tryptic digests followed by VVA LWAC enrichment and nLC-MS/MS is depicted in Fig. 1C (5). Tryptic digests from pairs of isogenic HEK293SC and HEK293SCAT with KO of individual GALNTs were differentially labeled with light (L) HEK293SC and medium (M) HEK293SCAT isotopes (Fig. 1C). We analyzed total cell lysates (TCL) and secretomes (SEC) separately, and global analysis of total data sets resulted in 3617 unambiguously assigned glycosites and 3289 (91%) of these were quantifiable (supplemental Fig. S3B). Quantification takes place at the MS1 level measuring relative abundance of the detected glycopeptide precursor ions, and this poses a number of complications with peptides containing more than one O-glycosites, apart from complexity in analysis. Interpretation of the contribution of individual GalNAc-Ts and their substrate specificities is ambiguous because GalNAc-Ts often cooperate in glycosylation of clustered glycosites (1). Most of the identified O-glycosites in the protease digests were in fact monoglycosylated peptides (supplemental Fig. S3C), in agreement with findings in our previous studies (5, 25). We therefore initially focused our analysis on glycopeptides with one GalNAc residue for the
Deconstruction of O-glycosylation in HEK293 Cells

**Fig. 1.** Graphic depiction of the gene engineering strategy and glycoproteomics workflow. A, Evaluation of expressed repertoire of GALNTs in HEK293 cells assessed by RNAseq analysis (top) and enzyme protein expression by ICC with mAbs (below). B, Depiction of the biosynthetic pathway of core1 O-glycosylation and the genes targeted for generation of isogenic HEK293SC and HEK293WT with KO of GALNT1, T2, T3, T7, T10, or T11. C, Depiction of the quantitative glycoproteomics workflow using dimethyl labeling. Trypsin digests from total cell lysates of paired isogenic HEK293SC and HEK293SCAT were labeled with “light” (L) and “medium” (M) isotopes, respectively, mixed and GalNAc-glycopeptides enriched on a long VVA column, and Labeled glycopeptides were analyzed by nLC-MS/MS.
comparative analysis of HEK293SC with KO of GALNT1, T2, T3, and T11.

The GALNT7 and T10 genes encode GalNAc-T isoforms with GalNAc-glycopeptide specificities, and they are believed to function as follow-up enzymes required to complete O-glycosylation of high-density regions (26–28). We therefore analyzed glycopeptides with both single and multiple glycosites for the GALNT7 and T10 data sets.

The O-glycoproteome Contributions of GALNT1, T2, and T3—The major contributors to the O-glycoproteome of cells are the GalNAc-T1 and T2 isoforms that have great overlap in substrate specificities and are rather ubiquitously expressed (5, 13). GalNAc-T3 also has great overlap in substrate specificity with GalNAc-T1 and T2, however its expression pattern is more restricted and markedly altered in cancer (1). Interestingly, GALNT3 is expressed in HEK293 but not in HEK293T cells widely used for recombinant expression of glycoproteins (unpublished information). Because these three isoforms were also among the highest expressed in our HEK293SC, we first focused on the differential glycoproteomes obtained with these cell lines (Fig. 2).

We analyzed histograms based on glycopeptide M/L ratios of all identified and quantified monoglycosylated peptides (glycopeptides with one GalNAc residue attached) (406, 412, and 465 for ΔT1, ΔT2, and ΔT3, respectively) for each of the paired HEK293SC/HEK293SCAT data sets (Fig. 2). Loss of either of the GalNAc-T1/T2/T3 isoforms is predicted to result in identification of loss of GalNAc-glycopeptides that are not covered by overlap and redundancy by other endogenous GalNAc-Ts with appearance of L singlets or low M/L ratios. However, glycopeptides with two O-glycosites of which only one is GalNAc-T isoform-specific can potentially result in detection of a glycopeptide with only one glycosite, which would appear as a M singlet based on the criteria of our analysis, which may explain the distribution with both light and medium shoulders and singlets.

The distribution of total quantified and unambiguously identified monoglycosylated peptides was similar among the three samples with similar relative numbers of unique glycopeptide identifications (Fig. 2D). We used a M/L cut-off ratio of 1:10 for selection of glycopeptides considered as candidates for GalNAc-T isoform-specific glycosylation, and included singlets in this analysis (Fig. 2A–2C), identifying a total of 7 glycopeptides unique for GalNAc-T1, 48 unique for GalNAc-T2, and 48 unique for GalNAc-T3 (Fig. 2E). We only found a few sites unique for GalNAc-T1 in contrast to T2 and T3, however, this may be related to low levels of expression of the close isoform GALNT13 (Fig. 1A), which has very similar activity (29). In our previous study in HepG2 cells we identified a considerably larger number of nonredundant GalNAc-T1 isoform-specific candidates (5), and expression of GALNT13 is not detectable by RNAseq. The unique sites identified for GalNAc-T2 and T3 were compared with our previous in vitro enzyme specificity analyses (13), and 7 of the unique glycosites identified for GalNAc-T2 and 6 for GalNAc-T3, corresponded to synthetic peptide substrates tested previously (supplemental Table S3). Most of these peptide substrates were exclusively or preferentially used by the corresponding isoenzymes, serving as one level of validation of the identified glycosites.

We also compared the identified unique glycosites to those previously identified in our study of differential glycoproteomes in HepG2 cell lines with KO of GALNT1 and T2, and knock in (KI) of GALNT3 (5). Fig. 3 presents a comparison of the obtained differential glycoproteomes for HepG2 cells and HEK293 cells. The two cell lines are of different origins and expected to have different proteomes, and to compare the different data sets, we extracted glycopeptides detected with quantification values in both HEK293 and HepG2 and plotted these (Fig. 3B, 3F, 3J). The overlap in data sets for GALNT1 was minimal, whereas the data sets for T2 showed that several glycopeptides were differentially down or upregulated in both, and always in the same direction (Fig. 3F). The same was true for the GALNT3 data sets when glycopeptides were plotted in opposite directions to account for the experimental difference with KI of GALNT3 in HepG2 cells and KO of GALNT3 in HEK293 (Fig. 3J).

The O-glycoproteome contributions of GALNT7 and T10—The GalNAc-T7 and T10 isoforms in subfamily Iib are classified as so-called follow-up enzymes with GalNAc-glycopeptide substrate specificities (1), and effects of loss of these isoforms are predicted to involve loss of one or more O-glycans on glycopeptides with multiple and clustered O-glycans. Analysis of the differential glycopeptide data with multiple GalNAc residues attached is challenging, not only analytically with respect to sensitivity and unambiguous assignments of the actual glycosites, but also regarding interpretation, because most glycopeptides with multiple occupied glycosites display inherent heterogeneity. Moreover, the major predicted substrates for these isoforms are the densely O-glycosylated tandem repeat regions of mucins, which are underrepresented in our data set because these regions are not digested with most common proteases, including trypsin used here. Nevertheless, we evaluated a number of metrics to better qualify the global effects on site occupancy that manipulation of follow-up enzymes may be expected to yield (Fig. 4). With an abrogation of activity of follow-up enzymes, quantitative data should reveal a relationship between the density of identified sites and the number of glycopeptides that are differentially identified. Specifically, when considering quantified glycopeptides, it would be expected that glycopeptides with >1 glycosites (i.e. glycopeptides that are potential substrates for follow-up glycosylation) comprise a higher proportion of the total differential glycopeptides compared with the background distributions of the total quantified glycopeptides (Fig. 4B). This pattern is clearly shown with the GALNT7 and GALNT7/GALNT10 knockout data. Here, we found a lower proportion of differential glycopeptides for 1xHexNAc
peptides compared with the background, whereas we found a higher proportion of glycopeptides displaying differential behavior for increasing densities from 3xHexNAc compared with the background. This distribution contrasts with that found for GALNT2 and GALNT3, where the differential glycopeptide density distribution follows the total quantified density distribution more closely. Unfortunately, the small number of unique quantified glycopeptides in the GALNT1, GALNT11...
and GALNT10 KO data set does not allow similar analysis to be performed.

We investigated methods to be able to identify differentially glycosylated glycosites from the complex differential data. To this end, we designed an algorithm that weighs abundance of evidence, when deducing the likely sites of differential regulation on glycopeptides. We produced a shortlist of regions that possibly harbor a differentially glycosylated substrate for GalNAc-T7 and T10. In contrast to singly glycosylated sites where singlet, or extreme quantification data is only used as evidence, inferences about follow-up glycosylation rely on the quantification of glycopeptides across the range of quantification ratios (Fig. 4A, 5). The current algorithm and supporting data does not support precise identification of differential sites, which means that we cannot summarize local amino acid sequence around sites for comparison with previous studies of follow-up glycosylation. However, this analysis does result in a shortlist for further investigation and validation.

The O-glycoproteome Contribution of GALNT11—Early in vitro analysis of GalNAc-T11 predicted a broad substrate specificity quite similar GalNAc-T1 and T2 (13, 23). More recent studies have shown that GalNAc-T11 may be the only isoform capable of glycosylating the short linker regions in LDLR class A (LA) repeats, although interestingly, GalNAc-T11 was unable in vitro to glycosylate peptides derived from the LA repeat regions (30, 31). The histogram based on glycopeptide M/L ratios of all identified single GalNAc-glycopeptides (356 for ΔT11) showed a tight Gaussian curve (Fig. 6), and remarkably few M singlets and more than 10x downregulated glycopeptides were identified in HEK293^SCAT11. Importantly, all the L singlets identified were from the linker regions of the LDLR related family including LRP2 and LRP8 (supplemental Table S2 and S3). We identified downregulated glycosites in the linker regions of LRP2 as well as in VLDLR and LRP1 (ambiguous sites identifications). In total, we identified 6 linker regions in VLDLR, LRP8, LRP1 and LRP2 with the consensus motif C^X^X^X^T^C^ (Fig. 6), and we predict that the glycosylation of all linker regions with this motif is specifically regulated by GalNAc-T11. We also identified glycosites in other glycoproteins (Fig. 6B) that were around 10-fold down-regulated. Interestingly, these were, in contrast to most of the identification in the LDLR related proteins, all very low intensities, and we anticipate that several of these may not be truly regulated by GalNAc-T11. In fact, none of these glycosites were identified in a preliminary analysis of the total cell lysates from the same HEK293 cell pair (not shown).

A Summary of the Global O-glycoproteome Identified in All Engineered HEK293^SC Cells—We previously analyzed the O-glycoproteome of the original HEK293^SC using a slightly different workflow and the less sensitive OrbiTrap XL instrument (25). Here, we also analyzed trypsin digests of HEK293^SC multiple times in the paired samples with individual GALNT mutant lines using the more sensitive Fusion instrument, and we greatly expanded the O-glycoproteome of HEK293 cells (supplemental Fig. 3A). We identified in total 2,767 O-glycosites, which represents an 8-fold expansion (supplemental Table S2).

DISCUSSION

Discerning unique functions of GalNAc-T isoenzymes is important to gain insight into how the large family of GalNAc-Ts orchestrate and differentially regulate the O-glycoproteome (1). Many of the GALNT genes cause or are predicted to cause or underlie susceptibilities to highly specific diseases and conditions (32, 33), and our knowledge of the molecular
Fig. 4. Summary of differential glycoproteomes of HEK293SC with KO of GALNT7, T10 or T7/T10. A, Upper Panel Histograms (gray) of log10 transformed quantification M/L ratios for all glycopeptides as presented in Fig. 2. The red overlaid histograms show the subset of quantified glycopeptides predicted to contain differentially glycosylated sites in HEK293 cells with KO of GALNT7, T10 or T7/T10. B, Histograms of the number of HexNAc residues on all quantified glycopeptides (gray, top) and the subset of differential (10-fold change) glycopeptides (green, bottom) are shown. For the KO of GALNT1/2/3/11 the number of identified glycopeptides peaks at 1 HexNAc, whereas for the KO of GALNT7/10 and T7/10 the number of glycopeptides with 10-fold changes (lower panel) peaks at 2–3 HexNAc in agreement with the predicted follow-up functions of these enzymes in high density O-glycosylated regions.
FIG. 5. Illustrative examples of differential glycopeptides identified in HEK293 cells with KO of GALNT7 or T7/T10. A, The stem region of LYPD6 is potentially specifically glycosylated by GalNAc-T7. Four glycopeptides are quantified in the differential glycoproteome. Glycopeptides with 2 and 3 HexNAc residues are down-regulated in GALNT7 KO cells, whereas a glycopeptide with a single site is upregulated, suggesting that there is less transformation from a single site peptide to a di or tri-glycosylated peptide in knockout cells. B, Similarly, in a GALNT7/T10 KO cell line, a highly glycosylated region of CD44 is potentially glycosylated by either GalNAc-T7 or GalNAc-T10. Glycopeptide species with single glycosylation sites are upregulated in GALNT7/T10 KO cell lines, whereas the species with 4 HexNAc residues is downregulated. In cells with a single knockout of GALNT7, no differentially glycosylated peptides are found in this region. Combined, this information suggests that either GalNAc-T10 glycosylates this region alone, or that it can work in concert with GalNAc-T7.

Fig. 5. Illustrative examples of differential glycopeptides identified in HEK293 cells with KO of GALNT7 or T7/T10. A, The stem region of LYPD6 is potentially specifically glycosylated by GalNAc-T7. Four glycopeptides are quantified in the differential glycoproteome. Glycopeptides with 2 and 3 HexNAc residues are down-regulated in GALNT7 KO cells, whereas a glycopeptide with a single site is upregulated, suggesting that there is less transformation from a single site peptide to a di or tri-glycosylated peptide in knockout cells. B, Similarly, in a GALNT7/T10 KO cell line, a highly glycosylated region of CD44 is potentially glycosylated by either GalNAc-T7 or GalNAc-T10. Glycopeptide species with single glycosylation sites are upregulated in GALNT7/T10 KO cell lines, whereas the species with 4 HexNAc residues is downregulated. In cells with a single knockout of GALNT7, no differentially glycosylated peptides are found in this region. Combined, this information suggests that either GalNAc-T10 glycosylates this region alone, or that it can work in concert with GalNAc-T7.
mechanisms behind these functions is still highly limited. Here, we developed two panels of isogenic HEK293 cells (SC and WT) with individual KO of all the major GALNT genes expressed, and used the SC panel to probe nonredundant contributions to the O-glycoproteome for six GalNAc-Ts. Differential O-glycoproteomics shows that each GalNAc-T provides unique contributions to the O-glycoproteome, and that these nonredundant contributions only represent a minor part of the O-glycoproteome, confirming that the majority of the O-glycoproteome is covered by redundancies among the most prevalent GalNAc-T isoenzymes. The most restricted unique contribution was found for the GalNAc-T11 isoform, where the nonredundant functions were essentially limited to a common motif (CxXXXTC) in the short linker regions of LDLR related receptors. These O-glycans in their sialylated forms markedly enhance the binding and uptake of LDL by LDLR and VLDLR (31). These findings stress the importance of identifying the unique substrates for individual GalNAc-T isoforms to explore their roles in health and disease.

The differential O-glycoproteomics strategy using VVA LWAC enrichment with pairs of isogenic SimpleCells differing in a single GALNT gene can provide deep insight into the nonredundant functions of individual GalNAc-T isoforms to explore their roles in health and disease. However, there are several caveats with the study design to consider when evaluating outcomes. The use of SimpleCells with truncated O-glycans could affect the outcome both in terms of use of glycosites and with respect to changes in the transcriptome/proteome of cells. Although the truncated O-glycans

---

**Fig. 6. Summary of differential glycoproteomes of HEK293SC with KO of GALNT11.** A, Histograms showing the distribution of M/L quantitation ratios of mono-glycopeptides identified. B, List of monoglycopeptides selectively lost in HEK293SC/H9004T11. C, Graphic depiction of O-glycosites in LA linkers of human LDLR-related receptors. Depicted are all identified O-glycosites (yellow squares) to date as well as predicted O-glycosites (open squares) in LA linker regions of the LDLR family proteins (31). Glycosites identified in this study shown to be regulated by GalNAc-T11 are shown in red squares.
provide improved sensitivity because the glycome is homogeneous and the VVA lectin is efficient in LWAC, we find that most of the glycosites identified in SCs are also found in WT cells, blood and/or organs indicating that the identified glycosites are normally used (25, 34, 35). KO of the COSMC gene in cell lines to produce SCs have been shown to produce subtle characteristic changes in the transcriptome (5, 20, 36), and it is possible that loss/gain of a glycosite reflects loss/gain of the underlying protein expression rather than difference in glycosylation capacity. This can partly be assessed by analyzing the proteome, but in many cases identification of peptides from low abundance O-glycoproteins have failed. Alternatively, for O-glycoproteins with multiple glycosites the identification of other glycosites may provide indicative expression of the protein. In cases where the same proteins are identified in different cell lines, validation can also be achieved by demonstrating dependence of GalNAc-T isoforms in multiple cell systems as illustrated in Fig. 3. Similarly, it is important to select a cell system where relevant substrate proteins are expressed, which has posed problems in the past and e.g. the important FGF23 substrate for GalNAc-T3 has not been identified in any SC line tested so far (37, 38). The strategy also does not report stoichiometries of glycosites, and glycosites with very low stoichiometry and/or on very low abundant O-glycoproteins may be prone to show variations that may be interpreted as regulated despite the 10-fold cutoff used. Regardless of these limitations, the strategy clearly provides relevant candidate lists for GalNAc-T isoform specific substrates (5, 20), as illustrated here for GalNAc-T11 by the identification of multiple regulated glycosites in the LDLR related proteins (Fig. 6).

Past in vitro studies of GalNAc-Ts predicted both overlap in functions as well as unique functions for individual isoforms (11), and the results obtained here with the differential glyco-proteomics strategy clearly recapitulates this scenario with the majority of the O-glycopeptide covered by redundancies. Only minor subsets of glycosites were distinctly controlled by the GalNAc-T11/T2/T3 isoforms and several of the glycosites were reproduced from our previous study in HepG2SC (Fig. 3) (5). We identified Roundabout homolog 1 (ROBO1) (Thr660 glycosite) as being a specific substrate for GalNAc-T2 in two independent studies using isogenic HepG2SC with KO of GALNT2 (5, 39). These findings agree with the distinct substrate phenotypes associated with deficiencies of these genes in animal models (1), and for GALNT2 and GALNT3 also in humans (34, 40).

The most surprising finding of the study was that the nonredundant contribution of the GalNAc-T11 isoform was almost exclusively limited to the linker regions of the LA modules in LDLR-related receptors. In past in vitro studies of this isoform we and others found only minor differences in the substrate specificity compared with GalNAc-T1, T2, and T3 (23, 41), and we only later showed that GalNAc-T11 selectively was able to glycosylate an E. coli expressed LDLR ectodomain in vitro (30). However, the present data suggests that GalNAc-T11 in fact primarily, if not exclusively, controls O-glycosylation of all the LDLR related receptors in the important ligand-binding region. In more recent studies, we have generated HEK293WT/ATT1 cell lines with inducible expression of GalNAc-T11, and a first differential analysis of O-glycosylation at different levels of enzyme expression and PNA lectin enrichment confirmed the ultra-specific function in glycosylating class A repeat linker sequences, and showed tight correlation of expression level of GalNAc-T11 with O-glycosylation of the linker sequences indicating the potential for regulation of stoichiometry on the LDLR related receptors (42). We recently demonstrated that the O-glycans in linker regions in LDLR and VLDLR strongly enhance ligand binding and uptake (31), and that these O-glycans are found in all the LDLR-related receptors, including LRP8 (ApoER) and LRP2 (Melin), we predict that the functions of all these endocytic receptors are regulated by GalNAc-T11. There are yet no reports of deficiency in the GALNT11 gene in rodents or man, but deficiency of the orthologous gene Pgant35A encoding dGalNAc-T1 in Drosophila results in developmentally lethal (41, 43). However, GALNT11 is a Genome-Wide-Association Study (GWAS) candidate gene for chronic kidney decline (44), and it is possible that dysregulation of this gene and resulting changes in O-glycosylation of LRP2 is involved.

Another important finding of the study was the confirmation that GalNAc-T7 and T10 as predicted from in vitro studies serve primarily as follow-up enzymes by incorporating GalNAc residues near already introduced GalNAc O-glycans (26, 28, 45). The catalytic domains of GalNAc-T7 and T10 are predicted to accommodate a GalNAc residue and glycosylate closely adjacent glycosites (11). Our analysis could only demonstrate that GalNAc-T7 and T10 unique sites were highly enriched for glycopeptides with high density of O-glycans. The nonredundant contributions of GalNAc-T7 appeared to be more substantial than T10 in the HEK293 cells, but the double KO of T7/T10 showed further loss of high-density glycopeptides compared with the individual KO suggesting that the two isoforms serve partly distinct functions as well. It is important to note that these studies were performed in HEK293 SimpleCells without detectable extensions of the initial GalNAc O-glycans and this may affect the follow-up functions of GalNAc-Ts. The lectin-mediated functions of GalNAc-Ts require binding to unmodified GalNAc residues in the close vicinity of acceptor sites, and the follow-up GalNAc-Ts (T4, T7, T10, T12) that recognize acceptor sites immediately adjacent to O-glycans only accommodate the initial GalNAc residue (11). It is thus not unlikely that lack of competition from elongation of O-glycans in SimpleCells may allow for more extensive elaboration of substrates by the GalNAc-Ts, and further studies with the HEK293 WT cells developed here are needed to explore these effects.

HEK293 cells are widely used for stable and transient expression of glycoproteins, and we developed HEK293 WT cell
lines with individual KO of the GALNT genes to enable further studies of the biological roles of the nonredundant O-glycans with the full glycan structure. The HEK293 WT/3T11 cell line has already been instrumental in demonstrating that the O-glycans in the class A linker regions of LDLR require sialic acid capping to enhance LDL binding and uptake (31), which may indicate that these negatively charged sialic acids add to the ionic interactions between Glu/Asp residues in the class A repeats and positive charged residues on lipoproteins and other ligands. Studies of other LDLR related receptors including LRPI and LRPII are now possible using these cells, and there are several reports indicating unique functions of different GALNT genes that can now be explored in greater detail in this panel of cells (46–49).

In summary, we present the hereto most comprehensive in cell analysis of functions of the GalNAc-T isoenzymes. The study confirms that individual isoenzymes serve limited and highly specific unique functions, whereas the major part of the O-glycoproteome is covered by functional redundancy. The engineered HEK293 WT and SC cell panels are useful for expression and analysis the regulation of O-glycoproteins and enable wider studies of the biological functions of GalNAc-T isofrom-specific glycosylation.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD009955.

* This work was supported by the Lundbeck Foundation, Novo Nordisk Foundation, Kirsten and Freddy Johansen Fonden, Læge Søfus Carl Emil Friis og hustru Olga Doris Friis’ Legat, and the Danish National Research Foundation (DNRF107).

† To whom correspondence may be addressed. E-mail: yoshi@sund.ku.dk.

‡ To whom correspondence may be addressed. E-mail: seva@sund.ku.dk.

Author contributions: Y.N., H.C., and S.Y.V. designed research; Y.N. and S.Y.V. performed research; Y.N., H.C., and S.Y.V. wrote the paper; J.H., A.H., Y.N. and S.Y.V. contributed new reagents/analytic tools.

REFERENCES

1. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22, 736–756.

2. Schjoldager, K. T., and Clausen, H. (2012) Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. Biochim. Biophys. Acta 1820, 2079–2094.

3. Gill, D. J., Thom, K. M., Chia, J., Wang, S. C., Steentoft, C., Clausen, H., Bard-Chapeau, E. A., and Bard, F. A. (2013) Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. Proc. Natl. Acad. Sci. U.S.A. 110, E3152–E3161.

4. Tran, D. T., and Ten Hagen, K. G. (2013) Mucin-type O-glycosylation during development. J. Biol. Chem. 288, 6921–6929.

5. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhruhev, S. Y., and Clausen, H. (2015) Deconstruction of O-glycosylation–GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. EMBO reports 16, 1713–1722.

6. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. J. Biol. Chem. 275, 38197–38205.

7. Revoredo, L., Wang, S., Bennett, E. P., Clausen, H., Moremen, K. W., Jarvis, D. L., Ten Hagen, K. G., Tabak, L. A., and Gerken, T. A. (2016) Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. Glycobiology 26, 360–376.

8. Gerken, T. A., Revoredo, L., Thome, J. J., Tabak, L. A., Vester-Chriestensen, M. B., Clausen, H., Galilay, G. K., Jarvis, D. L., Johnson, R. W., Moniz, H., and Moremen, K. W. (2016) Characterization of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-glycosylation. J. Biol. Chem. 288, 19900–19914.

9. de las Rivas, M., Lira-Navarrete, E., Daniel, E. J. P., Companon, I., Coelho, H., Diniz, A., Jimenez-Barbero, J., Peregrina, J. M., Clausen, H., Corza, D. Marcelo, F. Jimenez-Oses, G., Gerken, T. A., and Hurtado-Guerrero, R. (2017) The interdomain flexible linker of the polypeptide GalNAc transferase family dictates their long-range glycosylation preferences. Nat. Commun. 8, 1959.

10. Kubota, T., Shiba, T., Sugio, S., Furukawa, S., Sawaki, H., Kato, R., Wakisatsuki, S., and Narimatsu, H. (2006) Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (ppGalNAc-T10). J. Mol. Biol. 359, 708–727.

11. de las Rivas, M., Daniel, E. J. P., cloth, H., Lira-Navarrete, E., Raich, L., Companon, I., Diniz, A., Lagartera, L., Jimenez-Barbero, J., Clausen, H., Rovira, C., Rovira, A., Marcelo, F., Corzana, F., Gerken, T. A., and Hurtado-Guerrero, R. (2018) Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range Glycosylation Preferences of GalNAc-T4. ACS Cent. Sci., 4, 1274–1290.

12. Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moiinova, H., Ravi, L., Markowitz, S. D., Sher, A., and Tabak, L. A. (2011) Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. J. Biol. Chem. 286, 14493–14507.

13. Kong, Y., Joshi, H. J., Schjoldager, K. T., Madsen, T. D., Gerken, T. A., Vester-Chriestensen, M. B., Wandall, H. H., Bennett, E. P., Levery, S. B., Vakhruhev, S. Y., and Clausen, H. (2015) Probing polypeptide GalNAc transferase isoform substrate specificities by in vitro analysis. Glycobiology 25, 55–65.

14. Steentoft, C., Vakhruhev, S. Y., Vester-Chriestensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H. H., Levery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nucleic-acid-glycoengineered SimpleCell lines. Nat. Methods 8, 977–982.

15. Levery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., and Vakhruhev, S. Y. (2015) Advances in mass spectrometry driven O-glycoproteomics. Biochim. Biophys. Acta 1850, 33–42.

16. Steentoft, C., Bennett, E. P., and Clausen, H. (2013) Glycoengineering of human cell lines using zinc finger nucleic acid targeting: SimpleCells with homogeneous GalNAc O-glycosylation allow isolation of the O-glycoproteome by one-step lectin affinity chromatography. Methods Mol. Biol. 1022, 387–402.

17. Boersena, P. J., Foong, L. Y., Ding, V. M., Lemeer, S., van Breukelen, B., Philip, R., Boekhorst, J., Snel, B., den Hertog, J., Choo, A. B., and Heck, A. J. (2010) In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoadfinity purification and stable isotope dimethyl labeling. Mol. Cell Proteomics 9, 84–99.

18. Vakhruhev, S. Y., Steentoft, C., Vester-Chriestensen, M. B., Bennett, E. P., Clausen, H., and Levery, S. B. (2013) Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. Mol. Cell Proteomics 12, 932–944.

19. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalac-
tosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J. Biol. Chem.* 274, 25362–25370

20. Lavrsen, K., Dabelsteen, S., Vakhrushev, S. Y., Levery, S. B., Huan, A. D., Dylander, A., Mandel, U., Hansen, L., Frodin, M., Bennett, E. P., and Wandall, H. H. (2018) De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. *J. Biol. Chem.* 293, 1308–1314

21. Yang, Z., Steentoft, C., Hauge, C., Hansen, L., Thomesen, A. L., Niola, F., Vester-Christensen, M. B., Frodin, M., Clausen, H., Wandall, H. H., and Bennett, E. P. (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.* 43, e59

22. Mandel, U., Hassan, H., Therkildsen, M. H., Rygaard, J., Jakobsen, M. B., Juhi, B. B., Dabelsteen, E., and Clausen, H. (1999) Expression of polypeptide N-acetylgalactosaminyltransferase in a squamous cell carcinoma: immunohistochemical evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology* 9, 43–52

23. Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmim, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schafer, M. A., Haselmann, K., Zubarev, R., Roepstorff, P., Burchell, J. M., Taylor-Papadimitriou, J., Honour, W. M., and Clausen, H. (2002) Functional conservation of subfamilies of putative UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferases in Drosophila, Caenorhabditis elegans, and mammals. One subfamily composed of I(2)35Aa is essential in Drosophila. *J. Biol. Chem.* 277, 22623–22638

24. Vizzaino, J. A., Deutsch, E. W., Hassan, H., Hollingsworth, M. A., and Clausen, H. (1999) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv.* 38, 1408–1407

25. Steentoft, C., Vakhrushev, S. Y., Kong, Y., Vester-Christensen, M. B., Johansen, T. D., Seidah, N. G., Bennett, E. P., Levery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32, 1478–1488

26. Bennett, E. P., Hassan, H., Hollingsworth, M. A., and Clausen, H. (1999) A novel human UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. *FEBS Lett.* 460, 226–230

27. Chen, H., Tachibana, K., Zhai, Z., Sin, C., Kopp, K. L., Steentoft, C., Vakhrushev, S. Y., Olsen, J. V., Hansen, L., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Novel human UDP-N-acetylgalactosaminyltransferase isoforms to the O-glycoproteome by engineered SimpleCells.

28. Schjoldager, K. T., Vakhrushev, S. Y., Kong, Y., Stentoft, C., Vakhrushev, S. Y., Olsen, J. V., Hansen, L., Bennett, E. P., Woetmann, A., Yin, G., Chen, L., Song, H., Bakh, M., Hiday, R. A., Peters, S. L., Opavsky, R., Thode, C., Qvortrup, K., Schjoldager, K. T., Clausen, H., Hollingsworth, M. A., and Wandall, H. H. (2014) Immature truncated O-glycopeptide of cancer directly induces oncogenic features in human tumor cells. *Cell Death Differ.* 21, 767–778

29. Goth, C. K., Vakhrushev, S. Y., Joshi, H. J., Clausen, H., and Schjoldager, K. T. (2018) Fine-Tuning Limited Proteolysis: A Major Role for Regulated Site-Specific O-Glycosylation. *Trends Biochem. Sci.* 43, 269–284

30. Dzialowski, K., Jeanneau, C., Tarp, A. M., Benet-Pages, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., and Clausen, H. (2006) Polypeptide GalNAc-transferase 3 and familial tumoral calcinosis. *Science* 310, 1830–1837

31. Schjoldager, K. T., Vakhrushev, S. Y., Kong, Y., Steentoft, C., Nudelman, A. S., Pedersen, N. B., Wandall, H. H., Mandel, U., Bennett, E. P., Levery, S. B., and Clausen, H. (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycogeni-}
45. Perrine, C. L., Ganguli, A., Wu, P., Bertozzi, C. R., Fritz, T. A., Raman, J., Tabak, L. A., and Gerken, T. A. (2009) Glycopeptide-prefering polypeptide GaINAc transferase 10 (ppGaINAc T10), involved in mucin-type O-glycosylation, has a unique GaINAc-O-Ser/Thr-binding site in its catalytic domain not found in ppGalNAc T1 or T2. *J. Biol. Chem.* **284**, 20387–20397

46. Liu, F., Xu, K., Xu, Z., de Las Rivas, M., Wang, C., Li, X., Lu, J., Zhou, Y., Delso, I., Merino, P., Hurtado-Guerrero, R., Zhang, Y., and Wu, F. (2017) The small molecule luteolin inhibits N-acetyl-alpha-galactosaminyltransferases and reduces mucin-type O-glycosylation of amyloid precursor protein. *J. Biol. Chem.* **292**, 21304–21319

47. Schjoldager, K. T., Vester-Christensen, M. B., Bennett, E. P., Levery, S. B., Schwientek, T., Yin, W., Eliot, O., and Clausen, H. (2010) O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. *J. Biol. Chem.* **285**, 36293–36303

48. Tian, E., Hoffman, M. P., and Ten Hagen, K. G. (2012) O-glycosylation modulates integrin and FGF signalling by influencing the secretion of basement membrane components. *Nat. Commun.* **3**, 869

49. Wu, Y. M., Liu, C. H., Hu, R. H., Huang, M. J., Lee, P. H., Hsu, W. M., Huang, H. C., and Huang, M. C. (2011) Mucin glycosylating enzyme GALNT2 regulates the malignant character of hepatocellular carcinoma by modifying the EGF receptor. *Cancer Res.* **71**, 7270–7279