



# Ultradeep O-GlcNAc proteomics reveals widespread O-GlcNAcylation on tyrosine residues of proteins

Chunyan Hou<sup>a</sup> , Jingtao Deng<sup>a</sup>, Ci Wu<sup>a</sup>, Jing Zhang<sup>b</sup> , Stephen Byers<sup>a</sup>, Kelley W. Moremen<sup>c</sup> , Huadong Pei<sup>a</sup>, and Junfeng Ma<sup>a,1</sup>

Edited by Chi-Huey Wong, The Scripps Research Institute, La Jolla, CA; received May 12, 2024; accepted October 4, 2024

As a unique type of glycosylation, O-linked  $\beta$ -N-acetylglucosamine (O-GlcNAc) modification (O-GlcNAcylation) on Ser/Thr residues of proteins was discovered 40 y ago. O-GlcNAcylation is catalyzed by two enzymes: O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA), which add and remove O-GlcNAc, respectively. O-GlcNAcylation is an essential glycosylation that regulates the functions of many proteins in virtually all cellular processes. However, deep and site-specific characterization of O-GlcNAcylated proteins remains a challenge. We developed an ultradeep O-GlcNAc proteomics workflow by integrating digestion with multiple proteases, two mass spectrometric approaches (i.e., electron-transfer/higher-energy collision dissociation [ET<sub>h</sub>CD] and HCD product-dependent electron-transfer/higher-energy collision dissociation [HCD-pd-ET<sub>h</sub>CD]), and two data analysis tools (i.e., MaxQuant and Proteome Discoverer). The performance of this strategy was benchmarked by the analysis of whole lysates from PANC-1 (a pancreatic cancer cell line). In total, 2,831 O-GlcNAc sites were unambiguously identified, representing the largest O-GlcNAc dataset of an individual study reported so far. Unexpectedly, in addition to confirming known sites and identifying many other sites of Ser/Thr modification, O-GlcNAcylation was found on 121 tyrosine (Tyr) residues of 93 proteins. In vitro enzymatic assays showed that OGT catalyzes the transfer of O-GlcNAc onto Tyr residues of peptides and OGA catalyzes its removal. Taken together, our work reveals widespread O-GlcNAcylation on Tyr residues of proteins and that Tyr O-GlcNAcylation is mediated by OGT and OGA. As another form of glycosylation, Tyr O-GlcNAcylation is likely to have important regulatory roles.

glycosylation | O-GlcNAc | proteomics | tyrosine | OGT

Glycosylation is one of the most frequently occurring posttranslational modifications (PTMs) on proteins (1, 2). As a ubiquitous modification, protein glycosylation has critical roles in various biological events (3, 4). Among the structurally diverse glycans, O-linked  $\beta$ -N-acetylglucosamine (O-GlcNAc) modification (i.e., O-GlcNAcylation) was found to be a monosaccharide modification on Ser/Thr residues of intracellular proteins 40 y ago (5, 6). O-GlcNAc cycling on Ser/Thr residues is controlled by two enzymes. Using uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) as a substrate donor, O-GlcNAc transferase (OGT) adds O-GlcNAc to acceptor polypeptides (7, 8). O-GlcNAcase (OGA) hydrolyzes the glycosidic linkage to liberate free Ser/Thr residues (9). Given its essential regulatory roles, maintaining O-GlcNAcylation homeostasis is key to normal cellular physiology and deregulated protein O-GlcNAcylation underlies multiple human diseases (10–14). Targeting protein O-GlcNAcylation has shown great promise for biomedical applications (e.g., as therapeutic targets and biomarkers for cancer) (15).

Protein O-GlcNAcylation functions in a site-specific manner. Thus, determination of O-GlcNAc Ser/Thr sites of proteins generally serves as an initial step for functional characterization. To that end, great progress has been made in the past several decades, especially with the recent technological advances in high-throughput methods (e.g., mass spectrometry-based proteomics) (15–18). Much effort has been focused on the development of methods/materials that can effectively enrich O-GlcNAc peptides/proteins from complex samples prior to mass spectrometric analysis. By leveraging the (bio)chemical properties of the O-GlcNAc moiety, two categories of enrichment methods/materials have been developed. One is affinity enrichment, as exemplified by the use of antibodies (19–21), lectins (notably Wheat germ agglutinin) (22–24), phenylboronic acid (25), and by hydrophilic interaction chromatography (26, 27). Another approach is chemical/biochemical derivatization, including alkaline  $\beta$ -elimination followed by Michael addition (28, 29) and chemoenzymatic labeling followed by selective capture on solid supports (e.g., *via* click chemistry and other bioorthogonal reactions) (30–41). In contrast to in vitro chemoenzymatic labeling, metabolic sugar labeling (e.g., by using azido-derived analogs such as Ac<sub>4</sub>GalNAz and GalNAz) utilizes the

## Significance

Since its discovery 40 y ago, intracellular O-GlcNAcylation on Ser/Thr residues has become a well-known posttranslational modification with important functional roles. By using an integrated ultradeep O-GlcNAc proteomics workflow, we find that O-GlcNAcylation is present on over 120 tyrosine (Tyr) residues on proteins, in addition to confirming known sites and identifying many other sites of Ser/Thr modification. Moreover, we show that OGT catalyzes O-GlcNAcylation on Tyr, which can be removed by OGA. Collectively, we demonstrate that Tyr is a target for O-GlcNAcylation and that Tyr O-GlcNAcylation is another glycoform which can be mediated by OGT and OGA.

Author affiliations: <sup>a</sup>Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC 20007; <sup>b</sup>Department of Chemistry and Center for Diagnostics & Therapeutics, Georgia State University, Atlanta, GA 30302; and <sup>c</sup>Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602

Author contributions: C.H. and J.M. designed research; C.H., J.D., and C.W. performed research; J.Z., S.B., K.W.M., and H.P. contributed new reagents/analytic tools; C.H. and J.M. analyzed data; and C.H. and J.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

<sup>1</sup>To whom correspondence may be addressed. Email: junfeng.ma@georgetown.edu

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2409501121/-DCSupplemental>.

Published November 12, 2024.

UDP-GlcNAc salvage pathways in cells and thus simplifies the enrichment workflow (42–44). However, azido analogs should be carefully selected to avoid nonselective labeling and potential artifacts (45). In addition to effective methodology to isolate O-GlcNAc peptides, advanced mass spectrometric approaches have proven to be very helpful in revealing O-GlcNAc sites. Given the lability of the O-GlcNAc moiety, fragmentation methods derived from electron transfer dissociation (ETD), especially electron-transfer/higher-energy collision dissociation (EThcD) and higher-energy dissociation product ions-triggered EThcD (i.e., HCD-pd-EThcD), have shown great advantages for accurate site assignment (19, 46–48). These gradually evolving methods have facilitated site-specific O-GlcNAc proteomic analysis in multiple samples. So far, over 4,000 human proteins have been cataloged as O-GlcNAcylated with high confidence (49, 50). Despite these efforts, deep and large-scale O-GlcNAc site mapping has never been an easy task—it is still far from routine to unambiguously identify a few thousand O-GlcNAc sites in individual studies.

Here, we report the development of an ultradeep O-GlcNAc proteomics workflow by combining multiple-protease-based protein digestion, chemoenzymatic labeling-based enrichment, data acquisition by two mass spectrometric approaches (i.e., EThcD fragmentation and HCD-pd-EThcD fragmentation), and two easily adaptable data analysis tools (i.e., MaxQuant [MQ] and Proteome Discoverer [PD]). With this integrated strategy, we have identified 2,831 unambiguous O-GlcNAc sites from PANC-1 cells. Besides identifying many additional sites of Ser/Thr modification, we reveal that O-GlcNAcylation modifies 121 tyrosine (Tyr) residues on multiple proteins. Further *in vitro* enzymatic assays show that OGT is able to transfer O-GlcNAc onto Tyr residues of peptides and OGA can remove Tyr O-GlcNAcylation from peptides. Our work indicates that widespread Tyr O-GlcNAcylation may represent a unique glycoform and a dynamic regulatory mode of functional importance.

## Results

**Development of an Integrated Workflow for O-GlcNAc Proteomics.** Previous O-GlcNAc proteomics studies focused almost exclusively on specific aspects (particularly enrichment methods) for site-specific O-GlcNAc analysis. However, other aspects are not trivial. For example, trypsin, a popular enzyme that can generate peptides in the preferred mass range for modern mass spectrometers, has been predominately used in proteomics studies. Alternatively, the combinatorial use of other enzymes has the potential to increase the amino acid coverage and detection of modification sites of individual proteins and whole proteomes (51–56). Recent O-GlcNAc motif analysis suggests that many modification sites appear to be located in Ser/Thr-rich regions (containing few or none of the arginine/lysine residues targeted by trypsin) (57). Thus, trypsin digestion may produce unusually long peptides, making their identification, especially O-GlcNAc site determination, a challenging task. To that end, other enzymes [e.g., elastase (58) and AspN/chymotrypsin (59)] have been exploited for O-GlcNAc site mapping on several proteins. Given that trypsin was used for almost all reported workflows for O-GlcNAc proteomics, it is clear that the feasibility and performance of using multiple enzymes for large-scale O-GlcNAc site mapping is worth exploring. However, it is largely unknown whether the resulting peptides are amenable for hybrid-type ETD fragmentation methods (especially EThcD and HCD-pd-EThcD).

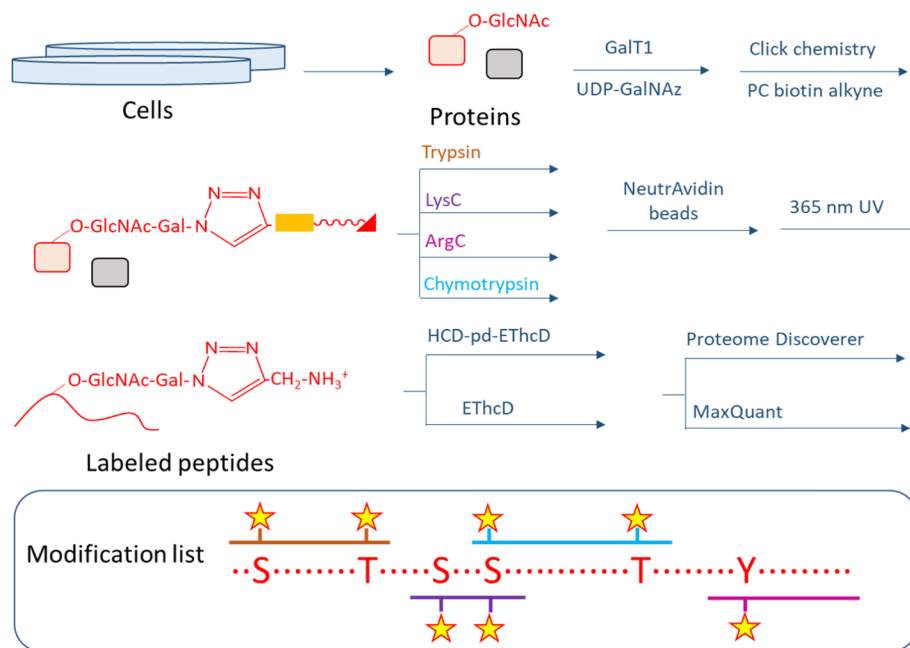
Our approach integrated multiple components, including sample preparation (i.e., chemoenzymatic labeling-based enrichment and multienzyme digestion), MS instrumentation (i.e., EThcD and HCD-pd-EThcD fragmentation modes), and two commonly used database searching software packages (i.e., MQ and PD) (Fig. 1). The integrated strategy for O-GlcNAc proteomics was benchmarked by analyzing cell lysates of PANC-1 (a pancreatic cancer cell line model). Four equal amounts of whole cell lysates (with 2 mg of total proteins in each) were subjected to chemoenzymatic labeling by using GalT1 (Y289L) and UDP-GalNAz (*SI Appendix, Fig. S1*). The resulting proteins were then reacted with a photocleavable (PC) biotin alkyne (which bears an ultra-violet [UV]-cleavable linker) via copper-based click chemistry. The tagged O-GlcNAc proteins were digested with four enzymes (i.e., trypsin, LysC, ArgC, and chymotrypsin) in parallel reactions. O-GlcNAc peptides in each digest were then captured on neutravidin beads and released after UV cleavage. The tagged O-GlcNAc peptides were analyzed by a nanoAcquity ultra-performance liquid chromatography (UPLC) (Waters) coupled with an Orbitrap Lumos mass spectrometer (Thermo Fisher Scientific) in either EThcD or HCD-pd-EThcD mode for two technical replicates. The resulting data were then analyzed by either MQ or PD. An overall > 50% enrichment selectivity was achieved in almost all the analyses (*SI Appendix, Fig. S2*). In total, more than 1.7 million MS/MS spectra were acquired, resulting in 2,544 unique O-GlcNAc peptides from 788 proteins (*SI Appendix, Table S1*). In total, 2,831 sites were identified with high confidence and as unambiguous O-GlcNAc sites (with a localization probability  $\geq 0.75$ ). A complete list of all the identified O-GlcNAc peptides, sites (including confidence scores and residues localized), and proteins for each enzyme, can be found in [Datasets S1–S3](#).

### O-GlcNAc Peptides/Sites Identified by the Integrated Workflow.

We evaluated the performance of each protease digestion for O-GlcNAc proteomics. Among the four datasets, trypsin provided the largest identification number of unique O-GlcNAc peptides (1,092), followed by chymotrypsin (767), ArgC (631), and LysC (569) (*SI Appendix, Table S1 and Fig. S3A*). Regarding unambiguous O-GlcNAc sites, trypsin yielded the highest identification number (1,336), followed by chymotrypsin (961), LysC (763), and ArgC (712) (*SI Appendix, Table S1 and Fig. S3B*).

We first investigated the global physicochemical characteristics of O-GlcNAc peptides (*SI Appendix, Fig. S4*). Although the length of O-GlcNAc peptides spans a wide range (6 to 91 amino acids), on average similar length distributions (16 to 32 amino acids) can be observed from the three enzymes (i.e., trypsin, ArgC, and LysC). Not surprisingly, identified O-GlcNAc peptides generated by chymotrypsin digestion are on average relatively shorter (13 to 23 amino acids). Correspondingly, O-GlcNAc peptide masses cross a wide range (i.e., 661 to 9,632 Da), with average masses between 1,940 to 3,650 Da for the three enzymes. Chymotryptic digestion produced relatively lower masses (1,595 to 2,760 Da) of peptides. Of note, it appears that distribution of mass to charge ( $m/z$ ) is quite similar among all the enzymes (mainly ranging from 611 to 952). Very strikingly, almost all O-GlcNAc peptide ions display over two positive charges (mostly 3 to 5 charges), which can be partially ascribed to the additional positive charge induced by the UV-cleavable tag. The triple and higher charges on O-GlcNAc peptides will undoubtedly facilitate their analysis by hybrid ETD fragmentation methods (i.e., EThcD and HCD-pd-EThcD).

We then evaluated the characteristics of O-GlcNAc sites identified from different proteases. Although trypsin identified the largest number of O-GlcNAc sites, significant improvement in



**Fig. 1.** An integrated workflow for ultradeep O-GlcNAc proteomics. The integrated system involves chemoenzymatic labeling-based enrichment, multiple-enzymatic digestion, two mass spectrometric approaches (i.e., ETHcD fragmentation and HCD-pd-ETHcD fragmentation), and two data analysis tools (i.e., MQ and PD).

O-GlcNAc coverage was detected with the use of other proteases (Fig. 2*A*), suggesting great complementarity of the enzymes. For example, by adding LysC to trypsin, the coverage of O-GlcNAc sites increased by 26%; and an additional increase of 30% was achieved by adding ArgC. More remarkably, the addition of three proteases (ArgC, LysC, and chymotrypsin) enhanced 112% of the distinctive O-GlcNAc sites (Fig. 2*B*). Collectively, the accumulated identification numbers of O-GlcNAc sites increased significantly, with an almost linear trend of increase with the use of four proteases. Among the O-GlcNAc sites, up to 73% (2,056) were not identified by more than one protease (Fig. 2*C*), indicating the high orthogonality of the multiprotease strategy.

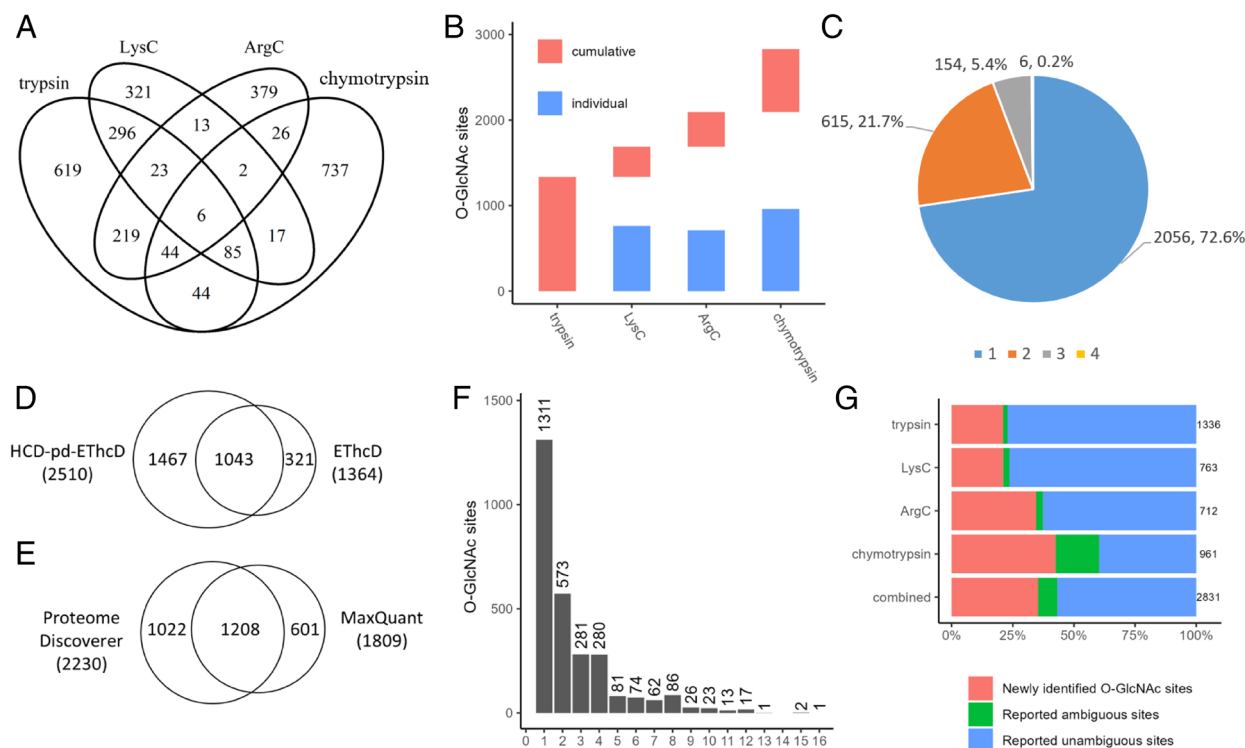
The performance of two mass spectrometry fragmentation methods ETHcD and HCD-pd-ETHcD (using HexNAc oxonium ions) was investigated for digests from each protease. In comparison to ETHcD, HCD-pd-ETHcD yielded ~1.5-fold (for chymotrypsin) to 3.5-fold (for other proteases) higher numbers of HexNAc-containing peptide-spectrum matches (*SI Appendix, Table S1*). Overall, 1.6-fold more O-GlcNAc peptides and ~1.8-fold more O-GlcNAc sites were identified by HCD-pd-ETHcD (*SI Appendix, Fig. S3 A and B*). Although HCD-pd-ETHcD accounted for ~88.7% (2,510) of all O-GlcNAc sites mapped, ETHcD contributed ~11.3% of additional O-GlcNAc sites (Fig. 2*D*). The combined usage of both dissociation approaches enables higher coverage of O-GlcNAc proteomics.

MQ (60) and PD, two software tools that render customization (e.g., allowing addition of the PC tag onto Ser/Thr/Tyr residues), were used for O-GlcNAc proteomics data analysis. The IMP-ptmRS algorithm built in PD (61) was enabled to facilitate confident localization of O-GlcNAc sites on the identified peptides. The O-GlcNAc sites with a localization probability over 0.75 in each software were considered as unambiguous sites. Similar numbers for O-GlcNAc identification across the two platforms were observed from each digest analyzed by either ETHcD or HCD-pd-ETHcD (*SI Appendix, Fig. S3 A and B*). However, these software tools appeared to identify a fairly distinct population of O-GlcNAc sites (with ~42.7% overlapped sites) (Fig. 2*E*). Given that each software program identifies unique O-GlcNAc

sites, the combination of both shows promise for further improvement of the coverage of O-GlcNAc sites.

Besides the complementarity, there is a remarkable coverage “redundancy” for specific O-GlcNAc sites. Overall, 775 O-GlcNAc sites were identified by at least two proteases (Fig. 2*C*), 1,043 sites were identified by both mass spectrometric approaches (Fig. 2*D*), and 1,208 sites were identified by both software tools (Fig. 2*E*). We then evaluated the identification frequency of each O-GlcNAc site by taking account of all approaches adopted. Including the use of four proteases, two fragmentation methods, and two data analysis packages, in total 16 different combinations were employed in our workflow (e.g., trypsin\_HCD-pd-ETHcD\_PD, trypsin\_ETHcD\_PD, trypsin\_HCD-pd-ETHcD\_MQ, trypsin\_ETHcD\_MQ, LysC\_HCD-pd-ETHcD\_PD, LysC\_ETHcD\_PD, and others). Among the 2,831 O-GlcNAc sites, 1,520 (~53.7%) sites were identified by two or more approaches (Fig. 2*F*). This prominent identification redundancy for O-GlcNAc sites clearly enhances data quality obtained from our integrated workflow. Next, we compared the O-GlcNAc sites identified from this work with several other studies/datasets. As O-GlcNAc levels vary substantially in different sample sources (57), we hypothesized that a more fair benchmark would be to compare against similar studies in which similar types of samples and/or experimental procedures were presented. The latest large-scale O-GlcNAc proteomic analysis on PANC-1 cancer cells was completed by us, in which trypsin was used as the sole enzyme, with O-GlcNAc peptides enriched by nitro-oxide-grafted nanospheres (via hydrophilic interaction) followed by analysis with HCD-pd-ETHcD mass spectrometry and PD (27). Even though a higher amount of starting material (5 mg of proteins) was used and high pH fractionation was applied, only 230 O-GlcNAc sites were unambiguously identified. Clearly the current tryptic-derived dataset yielded a much higher coverage of O-GlcNAc sites (*SI Appendix, Table S1*), which can be partially ascribed to the excellent enrichment efficiency of the chemoenzymatic-based labeling and tagging procedure. Our tryptic-derived dataset is comparable with another recent study which analyzed another cancer line (colorectal cancer cells) by using chemoenzymatic labeling-based enrichment, tryptic





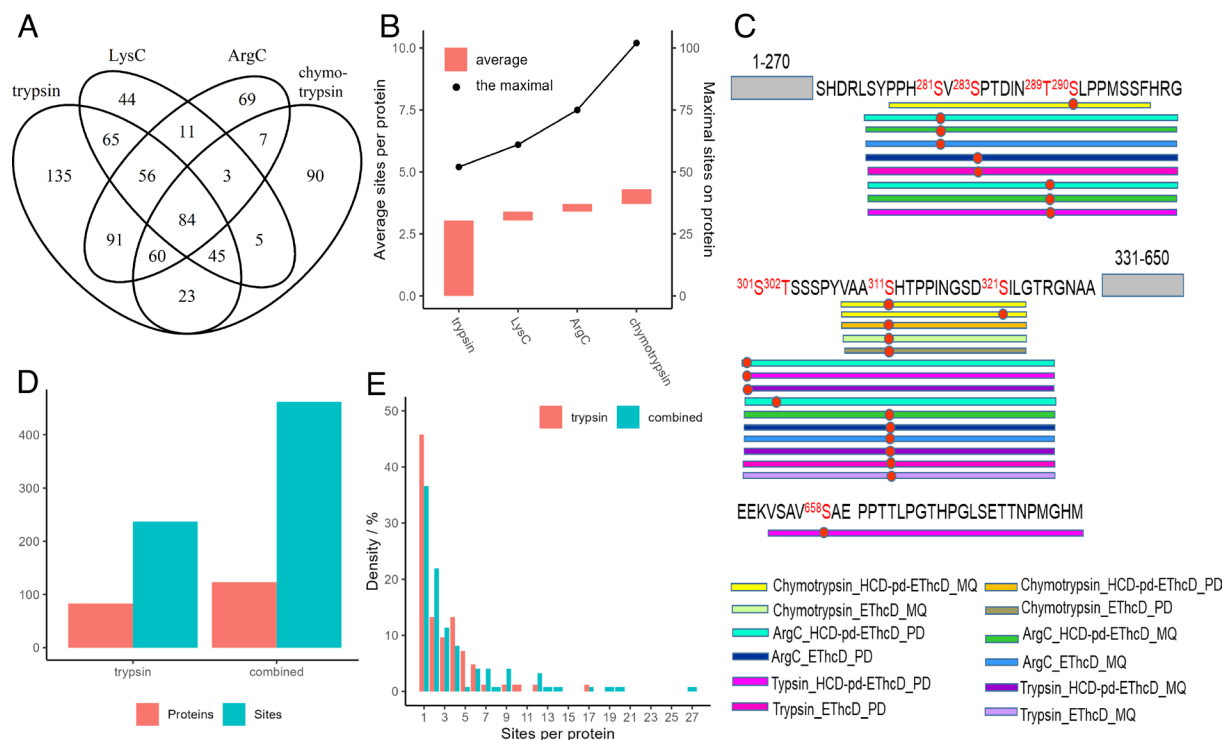
**Fig. 2.** A global view of O-GlcNAc sites identified from the integrated workflow. (A) Contribution of each protease. (B) Numbers of unique O-GlcNAc sites identified by each protease and the accumulation by each protease added. (C) The distribution of O-GlcNAc sites. The number such as “two” denotes the number of O-GlcNAc sites identified by two proteases. Overlap of unambiguous O-GlcNAc sites identified by using different approaches with different mass spectrometric methods (D) and different software tools (E). (F) Identification frequency of O-GlcNAc sites by different approaches used. The number such as two denotes the number of O-GlcNAc sites identified by two different approaches (e.g., Trypsin\_HCD-pd-ETHcD\_PD and trypsin\_HCD-pd-ETHcD\_MQ). (G) Overlapped O-GlcNAc sites between this work and previous dataset.

digestion, single-shot LC-MS/MS runs on HCD-pd-ETHcD mass spectrometry, and MQ (40). Nevertheless, it was clear that deeper O-GlcNAc proteomic coverage was achieved by using a workflow integrated with orthogonal proteases and other components. We then did another comparison against the datasets from O-GlcNAcAtlas 3.0 (49), which includes rigorously curated O-GlcNAc sites from both large-scale O-GlcNAc proteomics experiments and lower-throughput assays (e.g., based on immunoprecipitation with protein-specific antibodies followed by mass spectrometry) accumulating reported data from different human cells/tissue origins. We observed that our multiple-protease-based dataset not only confirms numerous unambiguous O-GlcNAc sites (1,607; accounting for 56%) on Ser/Thr residues reported before but also confidently identifies many sites (224, accounting for 8%) that were ambiguously assigned in previous studies (Fig. 2G). From the protease perspective, up to 79% of O-GlcNAc sites identified from tryptic digests and LysC digests were reported previously, while only 40% of O-GlcNAc sites from chymotryptic digests were reported. Moreover, chymotryptic digests appear to confidently identify a relatively high number of O-GlcNAc sites (18%) that were ambiguously assigned previously. We found many O-GlcNAc sites of Ser/Thr residues not yet reported. Although our trypsin dataset also yielded a number of additional sites (281), more than 71% (719) of the additional sites came from the datasets using the other proteases. Although trypsin performs very well and has been commonly used, it provides a somewhat biased representation of the O-GlcNAc proteome. This bias can be substantially reduced by utilizing the orthogonality of the alternative proteases shown here. Of particular interest, among the modification sites, O-GlcNAcylation was unambiguously found on 121 Tyr residues of a number of proteins.

Collectively, our dataset suggests that 1) there is a tryptic bias for large-scale profiling of O-GlcNAc sites, 2) there is high orthogonality between enzymes for O-GlcNAc analysis, and 3) the combined use of multiple proteases can substantially circumvent the tryptic bias, enabling high-coverage O-GlcNAc proteomics. Moreover, complementarity is observed between different mass spectrometric methods (ETHcD and HCD-pd-ETHcD) and different data analysis software tools (MQ and PD). Besides the significantly enhanced coverage, the identification of the majority of O-GlcNAc sites by two or more independent approaches suggests clear identification redundancy for O-GlcNAc sites obtained by the integrated workflow. Our data therefore reflect the value of integrating these complementary approaches, which allows deep and accurate O-GlcNAc proteomics.

#### O-GlcNAc Proteins Identified by the Integrated Workflow.

Besides distinct O-GlcNAc sites, different proteases increased the identification of O-GlcNAc proteins (Fig. 3A). Trypsin contributed the highest number (559) of O-GlcNAc proteins, while other proteases identified an additional 41% (229) O-GlcNAc proteins. In total, 145 O-GlcNAc proteins were newly identified (in comparison to the datasets from O-GlcNAcAtlas). Interestingly, despite the high number of unique O-GlcNAc peptides and unambiguous O-GlcNAc sites, chymotrypsin-digestion identified a relatively low number of O-GlcNAc proteins, as was the case with ArgC and LysC (Fig. 3A). These data also indicate that there might be increased O-GlcNAc density on proteins. Indeed, the average O-GlcNAc sites per protein was 3.0 by using trypsin, while the combined enzymes yielded an average of 3.7 O-GlcNAc sites per protein (Fig. 3B). Very strikingly, the maximum numbers of O-GlcNAc sites on a single protein were



**Fig. 3.** (A) Overlapping of O-GlcNAcylated proteins from each protease. (B) The identification and accumulation of average O-GlcNAc sites per protein from each protease (with the rising line indicating the maximum number of O-GlcNAc sites on a single protein). (C) Sequence illustration of identified peptides and corresponding O-GlcNAc sites on transcription factor 12 (TCF12) which contains 650 amino acids. Each frame below the sequence indicates the peptides identified, with the color representing one method used. O-GlcNAc sites are highlighted in red, with identified site(s) in the sequences marked as red dots. The gray frames show the parts that are not covered. (D) Histogram showing O-GlcNAcylated transcription factors (red) and O-GlcNAc sites (blue) by the traditional trypsin digestion only and the integrated methods using four proteases. (E) Distribution of O-GlcNAc sites on transcription factors.

sharply increased from 52 with tryptic digestion to 102 with four proteases (Fig. 3B).

Besides the O-GlcNAc density, we anticipated coverage redundancy for specific O-GlcNAc sites on certain proteins. Indeed, prominent coverage redundancy was observed for many O-GlcNAc sites on proteins (Dataset S3). For example, TCF12, a class-I E box-helix-loop-helix (bHLH) protein, is expressed in many tissues and can form heterodimers with other bHLH proteins (such as TWIST1) (62). Altogether, nine O-GlcNAc sites were identified on TCF12 (Fig. 3C). Among them, the method “trypsin\_HCD-pd-EThcD\_MQ” identified three sites (i.e., S281, S301, and S311), the method “trypsin\_HCD-pd-EThcD\_PD” identified three sites (i.e., T289, S301, and S658), the method “trypsin\_EThcD\_PD” identified two sites (i.e., S283 and S311), while the method “trypsin\_EThcD\_MQ” identified only one site (i.e., S311). Clearly, each method only covered a small portion of the O-GlcNAc sites on TCF12 while the combined usage of different methods would enhance the coverage. In total, tryptic digests (by coupling two fragmentation methods and two software tools) identified seven O-GlcNAc sites. As expected, the additional usage of other enzymes increased the number of O-GlcNAc sites identified. More importantly, the combinational approaches largely increased the identification frequency of specific O-GlcNAc sites. For example, O-GlcNAcylation on S311 was identified by up to 10 different approaches (Fig. 3C). The seemingly redundant identification for the same O-GlcNAc site is very important, given that the peptides <sup>275</sup>LSYPPHSVSPTDINTSLPPMSSFHR<sup>299</sup> and <sup>300</sup>GSTSSPYVAASHTPPINGS<sup>326</sup>SILGTR<sup>326</sup> are heavily O-GlcNAcylated (on four sites). It is well known that accurate and comprehensive O-GlcNAc site mapping of such peptides is notoriously challenging. Moreover, peptides such as these (with clustered Ser/Thr sites in the sequence) appear to be common

substrates for O-GlcNAcylation (as can be seen from the motifs). Undoubtedly, the application of the integrated strategy has advantages, which not only increases the coverage of O-GlcNAc sites but also allows unambiguous assignment of O-GlcNAc sites—the seemingly redundant identification strengthens the localization accuracy and assures high quality of the datasets.

Next, we performed a GO enrichment analysis of the 788 O-GlcNAc proteins. It turns out that highly enriched O-GlcNAc proteins are intracellular proteins in cytoplasmic granules, nuclear pores, the RNA-pol II transcription regulator complex, and histone acetyltransferase complex and other enzyme complexes (SI Appendix, Fig. S5A). In terms of molecular functions, enriched O-GlcNAc proteins are transcriptional regulators (coactivator or corepressor proteins that bind to and/or modulate the activity of transcription factors), proteins involved in nuclear receptor binding and histone binding, histone acetyltransferases, among others (SI Appendix, Fig. S5B). Biological process analysis reveals a broad spectrum of processes particularly chromatin organization and epigenetic regulation, including DNA/RNA transport, mRNA metabolism and stability, histone acetylation, and others (SI Appendix, Fig. S5C).

Indeed, transcriptional regulators are known to be a hot spot of protein O-GlcNAcylation. Emerging evidence shows that O-GlcNAc modifies many transcription factors, regulating their DNA binding, localization, stability, and interaction with cofactors (13, 57, 63, 64). But due to the low abundance of transcription factors and the substoichiometric nature of O-GlcNAcylation on proteins, analysis of O-GlcNAcylation on transcription factors on a proteome scale remains a challenge. Traditional O-GlcNAc proteomics studies using single trypsin digestion may mask the detection of functional O-GlcNAc sites, further hampering research on their biological roles. Thus, we anticipated that the

integrated workflow would allow large-scale detection of low-abundance O-GlcNAcylated transcription factors. Indeed, a total of 462 O-GlcNAc sites on 123 transcription factors annotated in the database of human transcription factors (65) were covered in our study (Fig. 3D; [Dataset S4](#)). Among them, 61% (282) of the O-GlcNAc sites were only reported here (rather than previous datasets). By comparing to the trypsin-only method, the identification of O-GlcNAcylated transcription factors and their O-GlcNAc sites was improved by 48% and 95%, respectively, with the integrated workflow (Fig. 3D). In contrast to the number of transcription factors, O-GlcNAc sites were more preferably recovered. Further examination unveiled that transcription factors with more than seven O-GlcNAc sites were better covered (Fig. 3E). Moreover, the average number of O-GlcNAc sites per transcription factor was increased from ~2.9 (237/83) to ~3.8 (462/123).

Given the substantially improved coverage and redundant identification of many O-GlcNAc sites, the integrated workflow will serve as a powerful strategy for in-depth and accurate O-GlcNAc proteomics for transcription factors or other subproteomes of interest. Remarkably, the significantly enhanced coverage and accuracy of O-GlcNAc sites identified on proteins allowed us to identify and characterize O-GlcNAcylation on Tyr residues.

**Characterization of Tyr O-GlcNAcylation.** The identification of O-GlcNAcylation on a relatively large number of Tyr residues is unprecedented. It not only reveals that this is a unique and widespread type of glycosylation but also provides us a glimpse into this previously unrecognized modification.

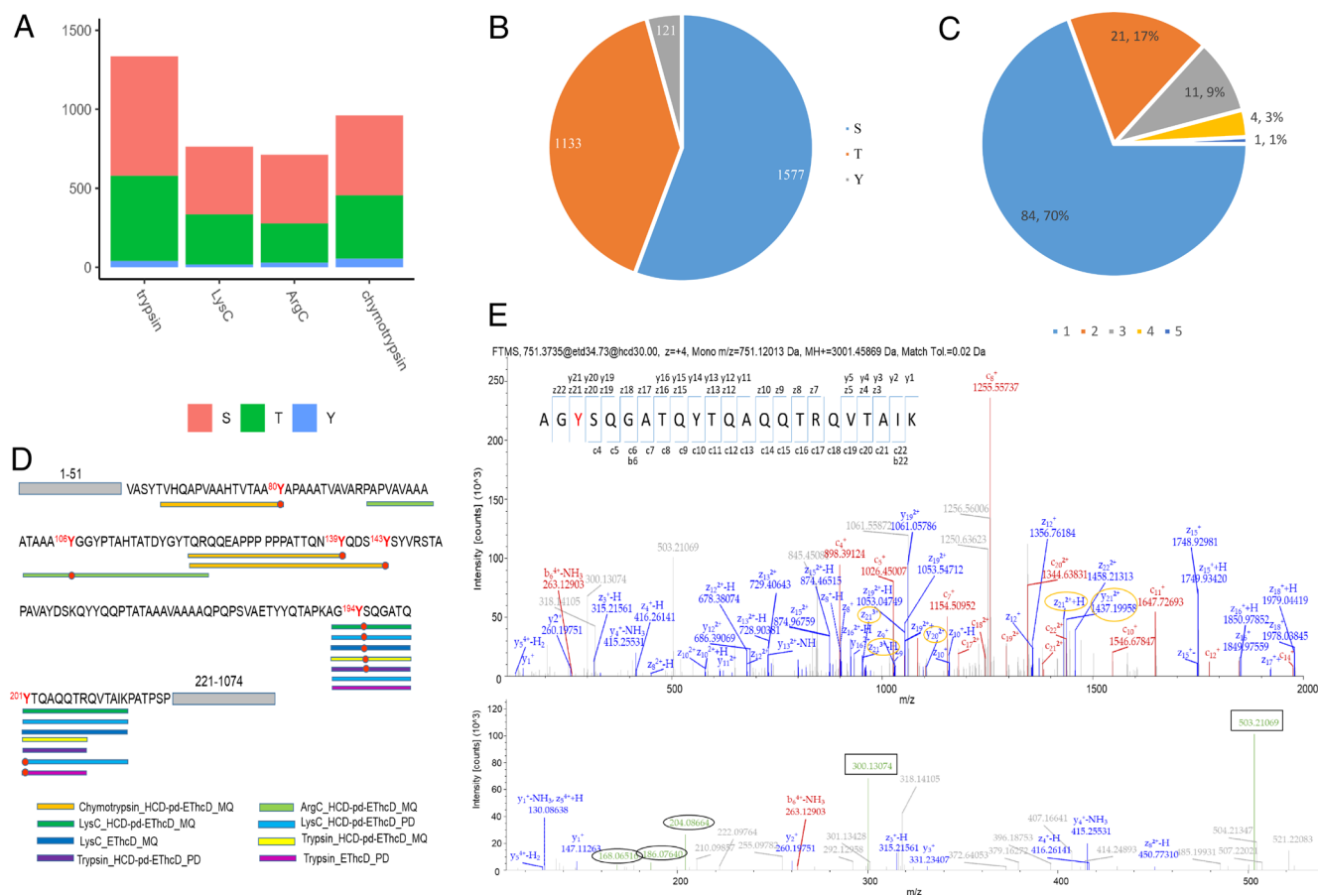
We calculated the numbers of O-GlcNAcylated Ser, Thr, and Tyr residues identified. Among the 1,336 O-GlcNAc sites, 40 Tyr O-GlcNAc sites (accounting for 3.0%) were identified from tryptic digests (Fig. 4A). Similar ratios of Tyr O-GlcNAcylation sites were obtained after either ArgC and LysC digestion. In contrast, out of 961 O-GlcNAc sites identified from chymotryptic digests, up to 5.6% (i.e., 54) sites were found on Tyr, suggesting that chymotrypsin may serve as a favorable enzyme to recover more Tyr O-GlcNAc sites. Cumulatively, the integrated workflow identified O-GlcNAcylation on 1,577 Ser sites, 1,133 Thr sites, and 121 Tyr sites, with a ratio of 56%:40%:4% (Ser:Thr:Tyr) (Fig. 4B). The ratio of Ser:Thr is somehow close to that reported in a previous meta-analysis report (in which Ser:Thr = 62%:38% was obtained for O-GlcNAc sites on human proteins) (57). Of note, “redundant” identification was also observed on Tyr O-GlcNAcylation sites, with 16% (37) of O-GlcNAcylated Tyr sites identified by two or more different approaches (Fig. 4C). For example, Zinc finger RNA-binding protein (ZFR) is a transcription factor involved in the regulation of growth and cancer development (66). The methods “trypsin\_HCD-pd-EThcD\_PD” and “trypsin\_HCD-pd-EThcD\_MQ” identified O-GlcNAcylation on Y194, while “trypsin\_EThcD\_PD” identified O-GlcNAcylation on Y201 of ZFR (Fig. 4D). In total, two O-GlcNAcylated Tyr sites (i.e., Y194 and Y201) were identified from the tryptic digests. In contrast, chymotrypsin and ArgC digestions yielded three additional sites (i.e., Y80, Y139, and Y143) and one site (Y106), respectively (Fig. 4D). Thus, the combination of all approaches unambiguously localized O-GlcNAcylation on six Tyr sites of ZFR (i.e., Y80, Y106, Y139, Y143, Y194, and Y201). Clearly, our strategy integrating multiple methods identified more O-GlcNAcylated Tyr sites than those that could be obtained by any single method. Even more importantly, the integrated strategy largely increased the identification frequency of specific Tyr O-GlcNAc sites. For example, O-GlcNAcylation on Y194 was identified by five different approaches (i.e., trypsin\_HCD-pd-EThcD\_PD, trypsin\_HCD-pd-

EThcD\_MQ, “LysC\_HCD-pd-EThcD\_MQ”, LysC\_HCD-pd-EThcD\_PD, and LysC\_EThcD\_MQ). A representative mass spectrum of O-GlcNAcylation on Y194 of the peptide <sup>192</sup>AGYSQGATQYTQAQQTRQVTAIK<sup>214</sup> from ZFR (with LysC digestion) is illustrated in Fig. 4E. As can be seen, the abundant presence of many fragment ions clearly shows that Y194 is O-GlcNAc modified, even though there are other Ser/Thr/Tyr sites in the peptide. In particular, the presence of discriminating fragments (i.e., c<sub>20</sub>, c<sub>21</sub>, y<sub>20</sub>, and y<sub>21</sub>) shows modification on Y194 rather than the neighboring S195. Moreover, O-GlcNAcylation on Y194 was also identified on a shorter peptide <sup>192</sup>AGYSQGATQYTQAQQTR<sup>208</sup> from ZFR (after trypsin digestion; with a representative mass spectrum shown in [SI Appendix, Fig. S6A](#)). Although only the HCD mass spectrum was acquired, the presence of abundant b-/y- ions enabled the assignment of O-GlcNAcylation on Y194 (with a localization score >0.99). For further confirmation, the O-GlcNAcylated form of the peptide “AGYSQGATQYTQAQQTR” (with modification on the first Tyr residue in its sequence) was synthesized and used for in vitro GalT1-based chemoenzymatic labeling followed by click chemistry (by using a similar procedure as for complex PANC-1 cell lysates). The resulting mass spectrum of the resulting O-GlcNAcylation peptide with a tag (after enrichment and UV cleavage) is shown in [SI Appendix, Fig. S6B](#). The almost identical mass fragmentation patterns confirm the O-GlcNAcylation on Y194. Of note, in a separate mass spectrum, we found that O-GlcNAc also modifies S195 on the same peptide (rather than the neighboring Y194; [SI Appendix, Fig. S6C](#)). Collectively, these data suggest that our integrated strategy not only enhances proteome coverage of O-GlcNAcylated Tyr sites but also provides redundant identification of many O-GlcNAcylated Tyr sites (enabling accurate localization of modification sites).

To explore the potential coexistence of O-GlcNAcylation on Tyr and Ser/Thr residues, we analyzed glycopeptides based on distinct stripped sequences [Datasets S1](#) and [S2](#). Besides 2,389 distinct stripped sequences with single modification, there are 465 sequences containing two or more sites (including 387, 75, and 3 sequences containing 2, 3, and 4 sites, respectively). Therefore, ~16% of all glycopeptide sequences are modified at more than one residue. Of note, these peptides were confidently identified with O-GlcNAcylation on multiple residues, even though not all sites were unambiguously allocated by the software (for a portion of such peptides). Regarding all unambiguously assigned O-GlcNAc sites on peptides, coexistence of O-GlcNAc on adjacent Ser and Thr appears to be relatively common (mass spectra of two glycopeptides shown in [SI Appendix, Fig. S7 A and B](#)). However, almost no peptides were found to have O-GlcNAcylation on both Tyr and adjacent Ser/Thr residues (± 5 amino acids) from the current dataset, indicating that O-GlcNAcylation on Tyr and Ser/Thr might be either mutually exclusive or possible when the target residues are localized far apart.

We then explored the potential relationship between Tyr O-GlcNAcylation and Tyr phosphorylation. Although the cross-talk between O-GlcNAcylation and phosphorylation on Ser/Thr residues is relatively well understood, whether such a relationship exists on Tyr residues is unknown. To that end, we compared the O-GlcNAcylated Tyr sites and 39,370 phosphorylated Tyr sites on human proteins from PhosphoSitePlus (67). Strikingly, up to 34% (41 sites) of the O-GlcNAcylated Tyr sites are localized to a known phosphorylated Tyr site nearby (within 10 amino acids) (Fig. 5A). It turns out that a very high percentage (~30.6%) of the O-GlcNAcylated Tyr sites actually coincide with phosphorylated Tyr sites [Dataset S5](#). For example, among the six O-GlcNAcylated Tyr sites of ZFR, three sites (Y143, Y194, and Y201) also exist in a phosphorylated form, and another site (Y139) is adjacent to a





**Fig. 4.** Characterization of O-GlcNAcylated Tyr sites. (A) Contribution of each protease to the identification of O-GlcNAcylated Ser/Thr/Tyr sites. (B) Ratio distribution of O-GlcNAcylated Ser/Thr/Tyr sites. (C) Identification frequency of O-GlcNAcylated Tyr sites by different approaches used. The number such as two denotes the number of O-GlcNAcylated Tyr sites identified by two different approaches (e.g., Trypsin\_HCD-pd-EthcD\_PD and trypsin\_HCD-pd-EthcD\_MQ). (D) Sequence illustration of identified peptides and corresponding Tyr O-GlcNAc sites on ZFR which contains 1,074 amino acids. Each frame below the sequence indicates the peptides identified, with the color representing one method used. O-GlcNAc sites are highlighted in red, with identified Tyr sites in the sequences marked as red dots. The gray frames show the parts that are not covered. Of note, O-GlcNAcylated Ser/Thr sites on ZFR are not displayed due to the space limit. (E) Representative mass spectrum of O-GlcNAcylation on Y194 of the peptide <sup>192</sup>AGYSQGATQYTQAQQTRQVTAIK<sup>214</sup> from ZFR (with LysC digestion) from the PANC-1 cell lysate. Of note, major fragments of the tag (i.e., m/z 503.21 and 300.13) are highlighted in black rectangles, and the HexNAc oxonium ions are shown in black circles in the *Bottom* panel. The key site-determining ions are highlighted with yellow circles.

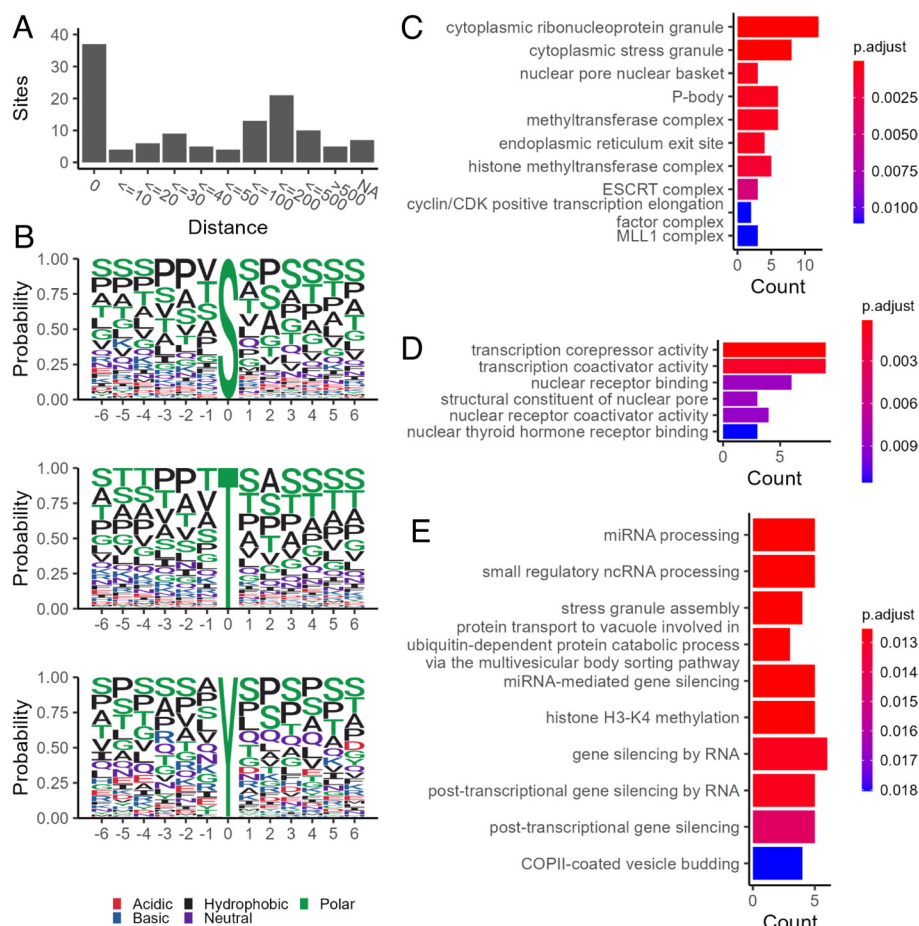
phosphorylated Tyr site (Y143). O-GlcNAcylation was also found on Y171 of the LIM and SH3 domain protein 1 (LASP1), which can be phosphorylated as well. It has been reported that phosphorylation on Y171 in the linker region of LASP1 abrogates its interaction with CXCR4 and AKT1, modulating diverse cellular processes (such as proliferation and cell survival) (68). Whether O-GlcNAcylation on Tyr sites on these and other proteins is involved in their functional regulation is worth exploring. Collectively, the high degree of colocalization suggests that there might be cross talk between these two PTMs (e.g., competing to occupy the same or adjacent Tyr sites). Although the colocalization analysis provides valuable insights into the potential functional roles, detailed studies will be needed to interrogate the cross-talk between the two PTMs.

Given that sequence analysis would provide a view about its potential modification pattern, we performed motif analysis for the Tyr O-GlcNAcylated peptides, together with Ser/Thr O-GlcNAcylated peptides (as a comparison) (Fig. 5B). It appears that a series of Ser residues flank a single O-GlcNAcylated Tyr site. Other than Ser residues, Pro residues are the most prevalent amino acid residues around the modification sites, scanning positions -6 (N-terminal from Tyr sites) through +6 (C-terminal from the Tyr sites). Moreover, Ala (a small amino acid) seems to be prominent as well. These features suggest similarities between Tyr

O-GlcNAcylation and Ser/Thr O-GlcNAcylation motifs, although slight differences can be observed. With the further increase of datasets of O-GlcNAcylated Tyr sites, motifs for Tyr O-GlcNAcylation may be improved.

To get further insights into the potential functions that Tyr O-GlcNAcylation may play, we performed GO functional analysis of Tyr O-GlcNAcylated proteins. Tyr-O-GlcNAc proteins are highly enriched in cytoplasmic ribonucleoprotein/stress granules, and (histone) methyltransferase complexes (Fig. 5C). These proteins are engaged in transcriptional regulation as coactivators, corepressors, nuclear receptors, among others (Fig. 5D). These enriched Tyr-O-GlcNAcylated proteins are involved in a variety of biological processes including miRNA/ncRNA processing, histone H3K4 methylation, and gene silencing (Fig. 5E). Site-specific functional studies will facilitate elucidation of the roles of Tyr O-GlcNAcylation in these processes.

**OGT Adds O-GlcNAc onto Tyr Residues.** One of the key questions regarding Tyr O-GlcNAcylation is whether it is an enzyme-catalyzed reaction, if so what might the enzymes be? Given that OGT is the enzyme that transfers GlcNAc from UDP-GlcNAc to Ser/Thr residues, we proposed that OGT might be an enzyme that has the capacity to mediate Tyr O-GlcNAcylation. To test that out, we did a series of OGT enzymatic assays *in vitro*, using synthetic peptides.



**Fig. 5.** Characterization of O-GlcNAcylated Tyr-containing peptides and proteins. (A) overlap with phosphorylated Tyr sites on human proteins. (B) Motif analysis of O-GlcNAcylated Ser/Thr/Tyr sites. GO classification of O-GlcNAcylated Tyr-containing proteins according to cellular component (C), molecular function (D), and biological process (E).

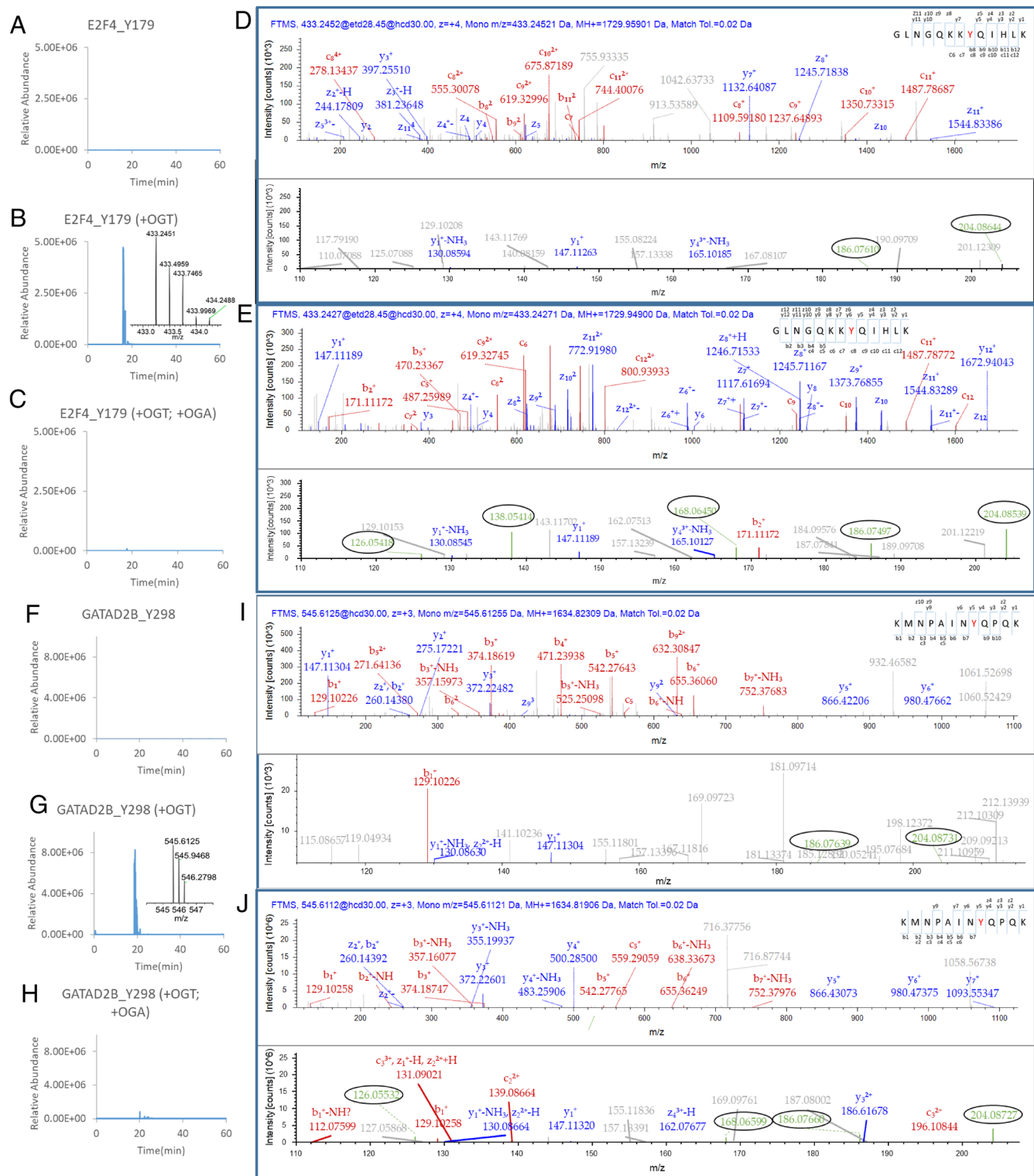
Peptide KM17 (“KKKYPGGSTPVSSANMM”) has been used an excellent substrate for OGT assays for over 10 y. Besides serving as a substrate to measure the enzyme activity of OGT, it is commonly used as a positive control for in vitro OGT reactions. When we analyzed the products of KM17 after incubating with OGT in vitro, we found that all the Ser and Thr residues (i.e., S8, T9, S12, and S13) were O-GlcNAcylated (SI Appendix, Fig. S8 A–D), confirming that it is indeed an excellent substrate for OGT. Unexpectedly, after careful examination of the mass spectra we found that Tyr residue (Y4) is also O-GlcNAcylated (SI Appendix, Fig. S8E). The presence of differentiating fragments (including  $c_3$ ,  $c_5$ ,  $z_{12}$ , and  $y_{14}$ ) clearly supports O-GlcNAcylation on Y4 of the peptide after the in vitro reaction catalyzed by OGT. Although the majority of glycopeptide species were modified on individual S/T/Y sites, it appeared the O-GlcNAcylation could occur on two residues simultaneously (e.g., Y4 and T9) (SI Appendix, Fig. S8 F and G). These unexpected observations not only confirmed the existence of bona fide Tyr O-GlcNAcylation but also prompted us to propose that Tyr O-GlcNAcylation is an enzymatic reaction and OGT is able to transfer the GlcNAc moiety onto Tyr residues of substrate peptides.

Although KM17 may serve as a good Tyr-containing substrate for OGT assay, it has multiple Ser/Thr residues that can also undergo O-GlcNAcylation. To avoid potential interference from Ser/Thr residues, we decided to assay peptides that do not contain Ser/Thr residues. “GLNGQKKYQIHLK” (containing Y179 of the transcription factor E2F4) and “KMNPAINYQPQK” (containing Y298 of the transcriptional repressor p66-beta (GATAD2B)), for in vitro OGT assay.

In comparison to the unmodified form (without O-GlcNAc), the peptide GLNGQKKYQIHLK became O-GlcNAcylated on the Tyr site after the OGT assay (with the extracted UPLC traces of the O-GlcNAcylated form ( $m/z = 443.2453$ ) before and after the OGT assay shown in Fig. 6 A and 6 B, respectively). One representative mass spectrum of the O-GlcNAcylated GLNGQKKYQIHLK is shown in Fig. 6D (with oxonium ions from the O-GlcNAc moiety shown in the Bottom panel). For further confirmation, an O-GlcNAc peptide GLNGQKKYQIHLK (in which Tyr was O-GlcNAcylated) was synthesized in our lab and analyzed by the same mass spectrometry system. As can be seen from Fig. 6E, the fragments from the synthetic O-GlcNAc peptide are well matched to those from the peptide after in vitro OGT assay. Interestingly, when a shorter form of the peptide “KYQIHLK” (covering Y179 of E2F4) was used for in vitro OGT assay, the Tyr residue was found to be O-GlcNAcylated as well (SI Appendix, Fig. S9A), with a similar fragmentation pattern as the synthetic O-GlcNAc peptide KYQIHLK (in which Tyr was O-GlcNAcylated; SI Appendix, Fig. S9B). These data indicate that even a short peptide KYQIHLK with only seven residues could be recognized by OGT for O-GlcNAcylation.

Another peptide KMNPAINYQPQK (containing Y298 of GATAD2B) was also O-GlcNAcylated after in vitro OGT assay. The extracted LC traces of the O-GlcNAcylated form ( $m/z = 545.6117$ ) before and after the OGT assay are shown in Fig. 6 F and G, respectively. In comparison to the unmodified form, the peptide was O-GlcNAcylated on the Tyr residue after OGT assay. One representative mass spectrum of the O-GlcNAcylated KMNPAINYQPQK is shown in Fig. 6I (with oxonium ions





**Fig. 6.** OGT transfers GlcNAc to Tyr residues on peptides and OGA removes it. UPLC traces of Tyr O-GlcNAcylated peptides GLNGQKKYQIHLK (from E2F4; A–C) and KMNPAINYQPQK (from GATAD2B; F–H) (N = 3). Insets in Fig. B and Fig. G show the 4+ and 3+ precursor ion masses of O-GlcNAcylated peptides GLNGQKKYQIHLK and KMNPAINYQPQK, respectively. Mass spectra of Tyr O-GlcNAcylated peptides GLNGQKKYQIHLK (D) and KMNPAINYQPQK (I) after in vitro OGT assay. Mass spectra of synthetic O-GlcNAc peptides GLNGQKKYQIHLK from E2F4 (E) and KMNPAINYQPQK from GATAD2B (J) in which Tyr was O-GlcNAcylated. The oxonium ions of HexNAc are circled out in the expanded panel below each mass spectrum (with rotated and dashed lines used to point out some annotations in the mass spectra if needed).

from the O-GlcNAc moiety shown in the *Bottom* panel). For further confirmation, an O-GlcNAc peptide KMNPAINYQPQK (in which Tyr was O-GlcNAcylated) was also synthesized and analyzed by the same mass spectrometry system. As can be seen from Fig. 6J, the fragments from the synthetic O-GlcNAc

peptide are well matched to those from the peptide after in vitro OGT assay.

Collectively, these results confirm that Tyr O-GlcNAcylation is an enzyme-catalyzed reaction and OGT is able to transfer the GlcNAc moiety onto Tyr residues of substrate peptides.

**OGA Removes Tyr O-GlcNAcylation.** It was very encouraging to know that OGT can catalyze Tyr O-GlcNAcylation. However, whether Tyr O-GlcNAcylation could be reversed, and if so by what enzyme(s), was unknown. Given that OGT and OGA add and remove O-GlcNAcylation on Ser/Thr, we hypothesized that OGA might be able to remove O-GlcNAc from Tyr residues. As a proof-of-concept, to the reaction product of KM17 (KKKYPGGSTPVSSANMM) after OGT assay, we performed denaturation followed by treatment with *Clostridium perfringens* NagJ (*CpNagJ*), a close homologue of human OGA (69). We found that *CpNagJ* treatment not only removed O-GlcNAcylation on Ser and Thr residues (i.e., S8, T9, S12, and S13) but also removed O-GlcNAcylation on the Tyr residue of KM17 (Y4), indicating that OGA might be the enzyme that removes O-GlcNAcylation from Tyr residues. We then performed *CpNagJ* treatment to the reaction products of GLNGQKKYQIHLK (containing Y179 of E2F4) and KMNPAINYQPQK (containing Y298 of GATAD2B) resulted from the in vitro OGT assay (with the extracted UPLC traces of the O-GlcNAcylated form of the two peptides shown in Fig. 6C and Fig. 6H, respectively). Again, *CpNagJ* was found to remove O-GlcNAcylation from Tyr residues on these peptides.

Taken together, our results confirm that Tyr O-GlcNAcylation is reversible and OGA is able to remove the GlcNAc moiety from Tyr residues.

## Discussion

Since its discovery 40 y ago, protein O-GlcNAcylation has been increasingly recognized as an essential PTM across almost all kingdoms of life (57). Numerous studies show that dysregulated O-GlcNAcylation in human cells is closely associated with multiple diseases. Despite significant progress, technically determining the full extent of site-specific O-GlcNAc modification of the proteome remains a challenge, hindering the elucidation of the functional roles and our understanding of this important PTM in general. Furthermore, O-GlcNAc on Ser/Thr residues has been thought to be the only way in which O-GlcNAcylation could affect biological processes in physiology and pathology.

Although significant technical advances (especially in enrichment methods and mass spectrometry techniques) have facilitated site-specific O-GlcNAc proteomic analysis, it is still far from routine to identify thousands of O-GlcNAcylation sites in individual studies. To address this issue, we first established a practical approach for ultradeep O-GlcNAc proteomics. To that end, we developed a strategy integrating all components in the analytical workflow, including multiple-protease-based protein digestion, chemoenzymatic labeling and click chemistry-based enrichment, data acquisition by two mass spectrometric approaches (i.e., EThcD fragmentation and HCD-pd-EThcD fragmentation), and two data analysis tools (i.e., MQ and PD). The performance of this highly integrated strategy for O-GlcNAc proteomics was benchmarked by analyzing cell lysates of PANC-1 (a pancreatic cancer cell line). Our data revealed that there is great complementarity between different proteolytic enzymes (i.e., trypsin, LysC, ArgC, and chymotrypsin) for digestion, mass spectrometric methods (EThcD and HCD-pd-EThcD) for data acquisition, and software tools (MQ and PD) for data analysis. For example, although trypsin digestion-based analysis performs well and is most commonly used, there is clear trypsin bias in the O-GlcNAc proteomics datasets reported previously. Given that almost all of the previous workflows exclusively used trypsin, this is a critical point that should be considered in future studies. The combined use of multiple proteases can substantially circumvent the tryptic bias, enabling deep and large-scale profiling of O-GlcNAc sites

(especially those that play regulatory roles). Besides the significantly enhanced coverage, our strategy yields redundant identification of numerous O-GlcNAc sites. Among the 2,831 O-GlcNAc sites unambiguously identified, > 53% were identified by two or more independent approaches. This seemingly redundant identification for O-GlcNAc sites clearly facilitates unambiguous assignment of O-GlcNAc sites, especially for those that clustered on peptides. Clearly, our integrated strategy affords high-coverage and accurate site localization of O-GlcNAc proteins. It is anticipated that this integrated strategy can be exploited for ultradeep and high-quality O-GlcNAc proteomics in other types of cells/tissues.

The datasets obtained from our integrated strategy offer a wealth of information regarding O-GlcNAc sites/peptides/proteins. Besides confirming numerous known O-GlcNAc sites, we found many additional sites of Ser/Thr modification. More importantly, it is with this unprecedented depth of O-GlcNAc proteomics that we were able to identify O-GlcNAcylation on 121 Tyr residues of 93 proteins. The revelation of Tyr O-GlcNAcylation will complement our understanding of protein O-GlcNAcylation. Even though many questions remain, the identified O-GlcNAcylated Tyr sites/peptides/proteins in the datasets shed some insights regarding this modification.

The overall ratio of O-GlcNAcylated Ser:Thr:Tyr is determined to be 56%:40%:4%, although slight changes in the percentage of O-GlcNAcylated Tyr could be observed by different proteases (3.0 to 5.6%). These results suggest that, Tyr O-GlcNAcylation is relatively rare compared to O-GlcNAcylation on Ser/Thr. However, the percentage of Tyr O-GlcNAcylation appears to be similar to that of Tyr phosphorylation in mammalian cells (which generally accounts for < 2%, 70). Motif analysis of the O-GlcNAcylated peptides shows quite similar patterns between Tyr O-GlcNAcylation and Ser/Thr O-GlcNAcylation, implying that O-GlcNAcylation on Tyr may share the same set of enzymes as Ser/Thr.

One basic question regarding Tyr O-GlcNAcylation is whether it is an enzyme-catalyzed reaction, if so what might be the enzyme(s). To elucidate this, several Tyr-containing peptides (without the presence of Ser/Thr residues) were tested for in vitro OGT assay. We found that OGT can transfer GlcNAc from UDP-GlcNAc to Tyr residues and OGA can remove O-GlcNAc from Tyr residues. These data illustrate that, Tyr O-GlcNAcylation is also an enzymatic reaction mediated by OGT and OGA (similar to Ser/Thr O-GlcNAcylation). This finding is of great importance, for several reasons.

First, it will facilitate characterization of the enzymology of OGT/OGA toward Tyr O-GlcNAcylation. Synthetic peptides are traditionally used to determine enzymatic kinetics of enzymes (7, 8, 71, 72). It is known that not all peptides are great substrates by OGT in vitro (7, 72). Although we have shown that several Tyr-containing peptides can serve as substrates for OGT, they may not be the best substrates. Providing that we have already generated a list of Tyr O-GlcNAcylated peptides, further screening of substrate peptides for OGT is worthwhile and ongoing. Considering that OGT has varied  $K_m$  values toward different substrates for Ser/Thr O-GlcNAcylation (72), we do not know whether this is true regarding Tyr O-GlcNAcylation. A detailed characterization of enzymatic kinetics will help provide a basic understanding of the cycling enzymes and the dynamic nature of Tyr O-GlcNAcylation. On the other hand, defining the precise relationship of Tyr O-GlcNAcylation to Ser/Thr O-GlcNAcylation in the context of proteome organization is certainly another aspect of interest for future studies.

Second, knowing the enzymes involved in Tyr-O-GlcNAcylation will allow exploration of the catalytic mechanisms. The elucidation

of the crystal structures of OGT and OGA has provided invaluable insights into their catalytic mechanisms and substrate recognition (73–77). Of note, OGT is a promiscuous enzyme. For example, besides catalyzing O-GlcNAcylation on Ser/Thr, OGT can transfer the GlcNAc moiety to Cys residues (78) and use UDP-glucose to add O-linked glucose onto proteins (79, 80). Thus, it is possible that the promiscuity of OGT also enables its recognition of relevant substrates for Tyr O-GlcNAcylation. The availability of a series of candidate substrate peptides/proteins provided in our resource will undoubtedly facilitate further investigation of the catalytic mechanisms of OGT/OGA (especially toward Tyr O-GlcNAcylation).

Third, our in vitro assays found that Tyr O-GlcNAcylation is a reversible modification, indicating a regulatory role. O-GlcNAcylation on Ser/Thr is essential in virtually all cellular processes examined. Whether Tyr O-GlcNAcylation plays a similarly essential and broad role is not known. However, our study hints at several potential roles.

The identified Tyr O-GlcNAcylated proteins are highly enriched in multiple cellular components (including cytoplasmic granules, enzyme complexes, and nuclear pores) and involved in a number of processes (e.g., nuclear receptor binding and transcriptional regulations). This indicates that Tyr O-GlcNAcylation may be another regulatory mechanism, similar as its analog Tyr phosphorylation (81). Considering the high occurrences of colocalization and adjacent localization between Tyr O-GlcNAcylation and Tyr phosphorylation sites, site-specific cross-talk with Tyr phosphorylation may serve as one way that Tyr O-GlcNAcylation could regulate protein functions. Interestingly, some nonreceptor Tyr kinases appear to require proline residues as part of their recognition motif (e.g., JAK2 has a preference for Pro at –2 and +3) (82), showing some similarity to OGT-catalyzed Tyr O-GlcNAcylation. The cross-talk with Tyr phosphorylation will provide another important but unique insight into the regulatory roles of Tyr O-GlcNAcylation. If true, their cross talk will certainly add another layer of complexity to the already complicated cross-talk network between Ser/Thr O-GlcNAcylation and phosphorylation. Although the numbers of Tyr O-GlcNAcylated proteins/sites identified are relatively small, their functional importance may not be as widespread as that of Ser/Thr O-GlcNAcylated proteins. Nevertheless, more Tyr O-GlcNAcylated proteins/sites will be identified by applying this and/or other methods for deep O-GlcNAc proteomics of other samples, and our understanding of the regulatory roles of Tyr O-GlcNAcylation will certainly increase.

In summary, we developed an integrated strategy for O-GlcNAc proteomics, enabling unprecedented depth and accurate localization of O-GlcNAc sites. Our datasets revealed the widespread presence of Tyr O-GlcNAcylation, another O-GlcNAcylation form on many proteins. We also found that OGT and OGA could add and remove O-GlcNAc onto/from Tyr residues, demonstrating that Tyr O-GlcNAcylation is an enzyme-catalyzed modification. Our finding of the reversible Tyr O-GlcNAcylation and the identification of Tyr O-GlcNAcylated sites on many proteins will serve as an important and key step to help interrogate Tyr site-specific functions. We anticipate that our work may open up a research area and stimulate more research interest in protein glycosylation, which will enhance our understanding of its roles in physiology and pathology.

## Materials and Methods

Total proteins from PANC-1 cells were extracted and enriched followed by digestion using different enzymes. Enriched O-GlcNAc peptides were subjected to nanoUPLC-MS/MS analysis by using a nanoAcquity UPLC system (Waters) coupled with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher). Resulting data files were processed with PD (PD, Thermo Fisher Scientific, version 2.4) with Sequest HT and MQ (MQ, Max-Planck-Institute of Biochemistry, version 2.3.0.0), followed by bioinformatic analysis. In vitro OGT(OGA) assays were performed by using synthetic peptides and purified OGT(CpNagJ). Tyr O-GlcNAcylated peptides were synthesized by using an adapted solid-phase peptide synthesis protocol. Detailed materials and methods are included in [SI Appendix](#) and [Datasets S1–S5](#).

**Data, Materials, and Software Availability.** The mass spectrometry data files have been deposited to the MassIVE, a public data repository (with the identifier: <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=43024c5871904dd-2b26240e73842f7d1>) (83). All other data are included in the manuscript and/or supporting information.

**ACKNOWLEDGMENTS.** This work was in part supported by NIH/NCI grant P30 CA051008, the Cancer Cell Biology Program Pilot Fund, and the GUMC institutional support. The Orbitrap Lumos Tribrid mass spectrometer was purchased with a gift from the Dekelbom Foundation. We appreciate Dr. Daan M. F. van Aalten for kindly providing CpNagJ and Dr. Matthew Pratt for kind suggestions on the synthesis of O-GlcNAc peptides. We thank Dr. Jun Yin for helpful discussions.

- G. W. Hart, R. J. Copeland, Glycomics hits the big time. *Cell* **143**, 672–676 (2010).
- K. W. Moremen, M. Tiemeyer, A. V. Nairn, Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).
- K. T. Schjoldager, Y. Narimatsu, H. J. Joshi, H. Clausen, Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).
- A. Varki et al., Eds., *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, ed. 4, 2022).
- C. R. Torres, G. W. Hart, Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259**, 3308–3317 (1984).
- G. D. Holt, G. W. Hart, The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. *J. Biol. Chem.* **261**, 8049–8057 (1986).
- R. S. Haltiwanger, G. D. Holt, G. W. Hart, Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:Peptide beta-N-acetylglucosaminyltransferase. *J. Biol. Chem.* **265**, 2563–2568 (1990).
- R. S. Haltiwanger, M. A. Blomberg, G. W. Hart, Glycosylation of nuclear and cytoplasmic proteins. purification and characterization of a uridine diphospho-N-acetylglucosamine:Polypeptide beta-N-acetylglucosaminyltransferase. *J. Biol. Chem.* **267**, 9005–9013 (1992).
- D. L. Dong, G. W. Hart, Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269**, 19321–19330 (1994).
- G. W. Hart, C. Slawson, G. Ramirez-Correa, O. Lagerlof, Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).
- M. R. Bond, J. A. Hanover, O-GlcNAc cycling: A link between metabolism and chronic disease. *Annu. Rev. Nutr.* **33**, 205–229 (2013).
- X. Yang, K. Qian, Protein O-GlcNAcylation: Emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **18**, 452–465 (2017).
- G. W. Hart, Nutrient regulation of signaling and transcription. *J. Biol. Chem.* **294**, 2211–2231 (2019).
- J. C. Chatham, J. Zhang, A. R. Wende, Role of O-Linked N-acetylglucosamine (O-GlcNAc) protein modification in cellular (patho)physiology. *Physiol. Rev.* (2020), 10.1152/physrev.00043.2019.
- J. Ma, C. Wu, G. W. Hart, Analytical and biochemical perspectives of protein O-GlcNAcylation. *Chem. Rev.* **121**, 1513–1581 (2021).
- N. M. Riley, C. R. Bertozzi, S. J. Pittner, A Pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Mol. Cell Proteomics* **20**, 100029 (2021).
- J. C. Maynard, R. J. Chalkley, Methods for enrichment and assignment of N-acetylglucosamine modification sites. *Mol. Cell Proteomics* **20**, 100031 (2021).
- M. E. Griffin, L. C. Hsieh-Wilson, Tools for mammalian glycoscience research. *Cell* **185**, 2657–2677 (2022).
- P. Zhao et al., Combining high-energy C-Trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* **10**, 4088–4104 (2011).
- A. Lee et al., Combined antibody/lectin enrichment identifies extensive changes in the O-GlcNAc sub-proteome upon oxidative stress. *J. Proteome Res.* **15**, 4318–4336 (2016).
- R. A. Burt et al., Novel antibodies for the simple and efficient enrichment of native O-GlcNAc modified peptides. *Mol. Cell Proteomics* **20**, 100167 (2021).
- J. C. Trinidad et al., Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol. Cell Proteomics* **11**, 215–229 (2012).
- A. K. Nagel, M. Schilling, S. Comte-Walters, M. N. Berkaw, L. E. Ball, Identification of O-linked N-acetylglucosamine (O-GlcNAc)-modified osteoblast proteins by electron transfer dissociation tandem mass spectrometry reveals proteins critical for bone formation. *Mol. Cell Proteomics* **12**, 945–955 (2013).
- S.-L. Xu et al., Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1536–E1543 (2017).
- X. Wang et al., A novel quantitative mass spectrometry platform for determining protein O-glycan dynamics. *Mol. Cell Proteomics* **15**, 2462–2475 (2016).
- B. Shen et al., A novel strategy for global mapping of O-GlcNAc proteins and peptides using selective enzymatic deglycosylation, HILIC enrichment and mass spectrometry identification. *Talanta* **169**, 195–202 (2017).



27. C. Wu *et al.*, Design and preparation of novel nitro-oxide-grafted nanospheres with enhanced hydrogen bonding interaction for O-GlcNAc analysis. *ACS Appl. Mater. Interfaces* **14**, 47482–47490 (2022).
28. L. Wells *et al.*, Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell Proteomics* **1**, 791–804 (2002).
29. J. Ma *et al.*, O-GlcNAcomic profiling identifies widespread O-Linked  $\beta$ -N-acetylglucosamine modification (O-GlcNAcylation) in oxidative phosphorylation system regulating cardiac mitochondrial function. *J. Biol. Chem.* **290**, 29141–29153 (2015).
30. P. M. Clark *et al.*, Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J. Am. Chem. Soc.* **130**, 11576–11577 (2008).
31. N. Khidkeel *et al.*, Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3**, 339–348 (2007).
32. Z. Wang *et al.*, Enrichment and site mapping of O-Linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol. Cell Proteomics* **9**, 153–160 (2010).
33. Z. Wang *et al.*, Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3**, ra2 (2010).
34. J. F. Alfaro *et al.*, Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7280–7285 (2012).
35. S. Wang *et al.*, Quantitative proteomics identifies altered O-GlcNAcylation of structural, synaptic and memory-associated proteins in Alzheimer's disease. *J. Pathol.* **243**, 78–88 (2017).
36. J. Li *et al.*, An isotope-coded photocleavable probe for quantitative profiling of protein O-GlcNAcylation. *ACS Chem. Biol.* **14**, 4–10 (2019).
37. J. Ma *et al.*, O-GlcNAc site mapping by using a combination of chemoenzymatic labeling, copper-free click chemistry, reductive cleavage, and electron-transfer dissociation mass spectrometry. *Anal. Chem.* **91**, 2620–2625 (2019).
38. V. N. Huynh *et al.*, Defining the dynamic regulation of O-GlcNAc proteome in the mouse cortex—the O-GlcNAcylation of synaptic and trafficking proteins related to neurodegenerative diseases. *Front. Aging* **2**, 757801 (2021).
39. S. Xu, F. Sun, R. Wu, A Chemoenzymatic method based on easily accessible enzymes for profiling protein O-GlcNAcylation. *Anal. Chem.* **92**, 9807–9814 (2020).
40. J. Liu *et al.*, Quantitative and site-specific chemoproteomic profiling of protein O-GlcNAcylation in the cell cycle. *ACS Chem. Biol.* **16**, 1917–1923 (2021).
41. Y. Chen *et al.*, Endo-M mediated chemoenzymatic approach enables reversible glycopeptide labeling for O-GlcNAcylation analysis. *Angew. Chem. Int. Ed. Engl.* **61**, e202117849 (2022).
42. D. J. Vocadlo, H. C. Hang, E.-J. Kim, J. A. Hanover, C. R. Bertozzi, A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9116–9121 (2003).
43. C. M. Woo *et al.*, Mapping and quantification of over 2000 O-linked glycopeptides in activated human t cells with isotope-targeted glycoproteomics (Isotag). *Mol. Cell Proteomics* **17**, 764–775 (2018).
44. X. Li *et al.*, Chemoproteomic profiling of O-GlcNAcylated proteins and identification of O-GlcNAc transferases in rice. *Plant Biotechnol. J.* **21**, 742–753 (2023).
45. W. Qin *et al.*, Artificial cysteine s-glycosylation induced by Per-O-acetylated unnatural monosaccharides during metabolic glycan labeling. *Angew. Chem. Int. Ed. Engl.* **57**, 1817–1820 (2018).
46. J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9528–9533 (2004).
47. R. J. Chalkley, A. Thalhammer, R. Schoepfer, A. L. Burlingame, Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8894–8899 (2009).
48. C. Singh, C. G. Zampronio, A. J. Creese, H. J. Cooper, Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of n-linked glycoproteins. *J. Proteome Res.* **11**, 4517–4525 (2012).
49. J. Ma, Y. Li, C. Hou, C. Wu, O-GlcNAcAtlas: A database of experimentally identified O-GlcNAc sites and proteins. *Glycobiology* **31**, 719–723 (2021), 10.1093/glycob/cwab003.
50. E. Wulff-Fuentes *et al.*, The human O-GlcNAcome database and meta-analysis. *Sci Data* **8**, 25 (2021).
51. R. H. Aebersold, J. Leavitt, R. A. Saavedra, L. E. Hood, S. B. Kent, Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6970–6974 (1987).
52. M. J. MacCoss *et al.*, Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7900–7905 (2002).
53. S. Gauci *et al.*, Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal. Chem.* **81**, 4493–4501 (2009).
54. P. Giansanti *et al.*, An augmented multiple-protease-based human phosphopeptide atlas. *Cell Rep.* **11**, 1834–1843 (2015).
55. R. M. Miller *et al.*, Improved protein inference from multiple protease bottom-up mass spectrometry data. *J. Proteome Res.* **18**, 3429–3438 (2019).
56. P. Sinitcyn *et al.*, Global detection of human variants and isoforms by deep proteome sequencing. *Nat. Biotechnol.* **41**, 1776–1786 (2023).
57. J. Ma, C. Hou, C. Wu, Demystifying the O-GlcNAc Code: A systems view. *Chem. Rev.* **122**, 15822–15864.
58. S. Hardivillé *et al.*, TATA-Box binding protein O-GlcNAcylation at T114 regulates formation of the B-TFIID complex and is critical for metabolic gene regulation. *Mol. Cell* **77**, 1143–1152e7 (2020).
59. D. N. Simon *et al.*, OGT (O-GlcNAc Transferase) Selectively modifies multiple residues unique to lamin A. *Cells* **7**, 44 (2018).
60. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
61. T. Taus *et al.*, Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **10**, 5354–5362 (2011).
62. V. P. Sharma *et al.*, Mutations of TCF12, encoding a basic-helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nat. Genet.* **45**, 304–307 (2013).
63. S. Ozcan, S. S. Andrali, J. E. L. Cantrell, Modulation of transcription factor function by O-GlcNAc modification. *Biochim. Biophys. Acta* **1799**, 353–364 (2010).
64. M. P. Parker, K. R. Peterson, C. Slawson, O-GlcNAcylation and O-GlcNAc cycling regulate gene transcription: Emerging roles in cancer. *Cancers (Basel)* **13**, 1666 (2021).
65. S. A. Lambert *et al.*, The human transcription factors. *Cell* **172**, 650–665 (2018).
66. H. Zhang, C. F. Zhang, R. Chen, Zinc finger RNA-binding protein promotes non-small-cell carcinoma growth and tumor metastasis by targeting the Notch signaling pathway. *Am. J. Cancer Res.* **7**, 1804–1819 (2017).
67. P. V. Hornbeck *et al.*, 15 years of PhosphoSitePlus®: Integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.* **47**, D433–D441 (2019).
68. E. Butt *et al.*, Phosphorylation-dependent differences in CXCR4-LASP1-AKT1 interaction between breast cancer and chronic myeloid leukemia. *Cells* **9**, 444 (2020).
69. F. V. Rao *et al.*, Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J.* **25**, 1569–1578 (2006).
70. J. V. Olsen *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
71. L. K. Kreppel, M. A. Blomberg, G. W. Hart, Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308–9315 (1997).
72. L. K. Kreppel, G. W. Hart, Regulation of a cytosolic and nuclear O-GlcNAc transferase role of the tetratricopeptide repeats. *J. Biol. Chem.* **274**, 32015–32022 (1999).
73. M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz, S. Walker, Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567 (2011).
74. N. L. Elsen *et al.*, Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* **13**, 613–615 (2017).
75. C. Roth *et al.*, Structural and functional insight into human O-GlcNAcase. *Nat. Chem. Biol.* **13**, 610–612 (2017).
76. B. Li, H. Li, L. Lu, J. Jiang, Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat. Struct. Mol. Biol.* **24**, 362–369 (2017).
77. C. M. Joiner, H. Li, J. Jiang, S. Walker, Structural characterization of the O-GlcNAc cycling enzymes: Insights into substrate recognition and catalytic mechanisms. *Curr. Opin. Struct. Biol.* **56**, 97–106 (2019).
78. J. C. Maynard, A. L. Burlingame, K. F. Medzhiradzky, Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), a new post-translational modification in mammals. *Mol. Cell Proteomics* **15**, 3405–3411 (2016).
79. D. L. Shen *et al.*, Catalytic promiscuity of O-GlcNAc transferase enables unexpected metabolic engineering of cytoplasmic proteins with 2-Azido-2-deoxy-glucose. *ACS Chem. Biol.* **12**, 206–213 (2017).
80. N. Darabedian, J. Gao, K. N. Chuh, C. M. Woo, M. R. Pratt, The metabolic chemical reporter 6-Azido-6-deoxy-glucose Further reveals the substrate promiscuity of O-GlcNAc transferase and catalyzes the discovery of intracellular protein modification by O-Glucose. *J. Am. Chem. Soc.* **140**, 7092–7100 (2018).
81. T. Hunter, Tyrosine phosphorylation: Thirty years and counting. *Curr. Opin. Cell Biol.* **21**, 140–146 (2009).
82. Y. Deng *et al.*, Global analysis of human nonreceptor tyrosine kinase specificity using high-density peptide microarrays. *J. Proteome Res.* **13**, 4339–4346 (2014).
83. C. Hou *et al.*, Ultradeep O-GlcNAc proteomics reveals widespread O-GlcNAcylation on tyrosine residues of proteins. *MassIVE*. <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=43024c5871904dd2b26240e73842f7d1>. Deposited 5 May 2024.