Chemoproteomic profiling of O-GlcNAcylated proteins and identification of O-GlcNAc transferases in rice

Xilong Li[1,†] , Cong Lei[2,3,†], Qitao Song[2,4], Lin Bai[1,5], Bo Cheng[2,3], Ke Qin[2,3], Xiang Li[2,3], Boyuan Ma[1,5], Bing Wang[1], Wen Zhou[2,3], Xing Chen[2,3,4,6,7,*] and Jiayang Li[1,5,*]

[1]*State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China*

[2]*College of Chemistry and Molecular Engineering, Peking University, Beijing, China*

[3]*Beijing National Laboratory for Molecular Sciences, Peking University, Beijing, China*

[4]*Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China*

[5]*University of Chinese Academy of Sciences, Beijing, China*

[6]*Synthetic and Functional Biomolecules Center, Peking University, Beijing, China*

[7]*Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, Peking University, Beijing, China*

Summary

O-linked β-*N*-acetylglucosaminylation (O-GlcNAcylation) is a ubiquitous post-translation modification occurring in both animals and plants. Thousands of proteins along with their O-GlcNAcylation sites have been identified in various animal systems, yet the O-GlcNAcylated proteomes in plants remain poorly understood. Here, we report a large-scale profiling of protein O-GlcNAcylation in a site-specific manner in rice. We first established the metabolic glycan labelling (MGL) strategy with *N*-azidoacetylgalactosamine (GalNAz) in rice seedlings, which enabled incorporation of azides as a bioorthogonal handle into O-GlcNAc. By conjugation of the azide-incorporated O-GlcNAc with alkyne-biotin containing a cleavable linker via click chemistry, O-GlcNAcylated proteins were selectively enriched for mass spectrometry (MS) analysis. A total of 1591 unambiguous O-GlcNAcylation sites distributed on 709 O-GlcNAcylated proteins were identified. Additionally, 102 O-GlcNAcylated proteins were identified with their O-GlcNAcylation sites located within serine/threonine-enriched peptides, causing ambiguous site assignment. The identified O-GlcNAcylated proteins are involved in multiple biological processes, such as transcription, translation and plant hormone signalling. Furthermore, we discovered two O-GlcNAc transferases (OsOGTs) in rice. By expressing OsOGTs in *Escherichia coli* and *Nicotiana benthamiana* leaves, we confirmed their OGT enzymatic activities and used them to validate the identified rice O-GlcNAcylated proteins. Our dataset provides a valuable resource for studying O-GlcNAc biology in rice, and the MGL method should facilitate the identification of O-GlcNAcylated proteins in various plants.

Introduction

Rice is one of the most important staple crops, feeding more than half of the world's population (Fukagawa and Ziska, 2019; Yu et al., 2021). Comprehensive understanding of rice biology is crucial for improving agricultural production in response to food pressure caused by human population growth (El Sayed et al., 2021) and global climate change (Schneider and Asch, 2020). Furthermore, as one of the major model crop plants, rice has been widely and extensively studied in molecular genetics and genomics for decades (Chen et al., 2022; Goff, 1999). By contrast, studying rice proteins and their post-translational modifications (PTMs) at the proteome level by proteomics remains limited. One of the main reasons is lacking powerful tools for labelling and enrichment of the modified proteins.

Numerous types of PTMs occur on proteins that greatly increase the complexity of proteome (Walsh et al., 2005). Post-translational modifications play key roles in regulating protein activity (Sasaki et al., 2011), subcellular localization (Salomon and

Orth, 2013) and protein–protein interaction (Li and Kohler, 2014). Of all PTMs, glycosylation is the most ubiquitous and complex modification (Schjoldager et al., 2020). It has been estimated that more than 50% of the human proteins are glycosylated (Wong, 2005). Protein glycosylation includes N-linked glycosylation, mucin-type O-linked glycosylation and O-GlcNAcylation (Pucci et al., 2021). Unlike the complex N-glycans and mucin-type O-glycans decorating cell surface and secreted proteins (Schjoldager et al., 2020), O-GlcNAcylation attaches a monosaccharide GlcNAc onto serine or threonine residues of various intracellular proteins (Yang and Qian, 2017). In mammalian cells, a pair of enzymes, O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA), catalyse the addition and removal of O-GlcNAc, making O-GlcNAcylation reversible and dynamic (Gao et al., 2001; Haltiwanger et al., 1990; Lubas et al., 1997). Several thousand proteins have been identified to be O-GlcNAcylated in mammalian cells (Ma et al., 2021). O-GlcNAcylation regulates a variety of biological processes, such as stress response, transcription and signal transduction (Bond and Hanover, 2015; Wells et al., 2001). Dysregulation of O-GlcNAcylation has been implicated in human

diseases, including cancer and neurodegenerative diseases (Bond and Hanover, 2013; Slawson and Hart, 2011; Yuzwa and Vocadlo, 2014).

O-GlcNAcylation also occurs in plants. Two putative OGTs, SPINDLY (SPY) and SECRET AGENT (SEC), were identified in *Arabidopsis thaliana* based on their sequence and structural homology to the animal OGT (Olszewski et al., 2010). While the enzymatic activity of OGT has been confirmed for recombinant SEC (Hartweck et al., 2002), the enzymatic activity of SPY was not clear until it was biochemically characterized to be in fact a protein O-fucosyltransferase (Zentella et al., 2017). Furthermore, whether there is a plant OGA remains an open question. On the contrary, identification of O-GlcNAcylated proteins in plants lags behind that in animals (Ma et al., 2021). Several *Arabidopsis* and wheat proteins were identified to be O-GlcNAcylated and functionally studied (Xiao et al., 2014; Zentella et al., 2016). Large-scale identification of O-GlcNAcylated proteins in plants has only been performed in two cases. By using lectin weak affinity chromatography (LWAC) to enrich O-GlcNAcylated peptides, 262 and 168 O-GlcNAcylated proteins were identified from *A. thaliana* flower tissues and winter wheat plumules, respectively (Xu et al., 2017, 2019). Many of the identified O-GlcNAcylated proteins are involved in plant-specific processes such as flower development and hormone response. Notably, only a few proteins were identified in both studies, suggesting that plant O-GlcNAcylation may have species and/or tissue specificity. It is therefore of great interest to explore O-GlcNAcylation in other plant species, especially the staple food crops such as rice.

Here, we report a large-scale identification of O-GlcNAcylated proteins and modification sites in rice seedlings by metabolic labelling of O-GlcNAc with azidosugars, which enabled chemoselective tagging of O-GlcNAc with a fluorophore or an affinity tag via click chemistry. By enrichment and mass spectrometry analysis of the tagged O-GlcNAcylated proteins, a total of 1591 unambiguous O-GlcNAcylation sites and 811 O-GlcNAcylated proteins were identified. The identified O-GlcNAcylated proteins are involved in important biological processes in rice such as transcription, translation and hormone signalling. Furthermore, we identified two rice OGTs, OsOGT1 and OsOGT2 and confirmed their enzymatic activities. By co-expressing OsOGT1/2 in *Escherichia coli* and *Nicotiana benthamiana* leaves, a serious of proteins from the identified O-GlcNAcylated proteome was validated to be *bona fide* O-GlcNAcylated proteins. Our results add a valuable resource for investigating O-GlcNAc biology in plants.

Results and discussion

Metabolic labelling of O-GlcNAc with azides in rice seedlings

In mammalian systems, the metabolic glycan labelling method has proven effective for profiling of protein O-GlcNAcylation (Liu et al., 2021; Qin et al., 2017, 2018, 2020; Woo et al., 2018; Woo and Bertozzi, 2016). *N*-azidoacetylgalactosamine (GalNAz) and *N*-azidoacetylglucosamine (GlcNAz), once inside cells, are metabolically converted to UDP-GalNAz and UDP-GlcNAz through the salvage pathways, respectively (Figure S1). The two UDP-*N*-azidoacetylhexosamine (UDP-HexNAz) donors are interconverted by UDP-galactose-epimerase, resulting in the incorporation of azides into various GalNAc- or/and GlcNAc-containing glycans, including N-linked glycans, mucin-type O-linked glycans and O-

GlcNAc. Although various types of glycoproteins are eventually identified, O-GlcNAcylated proteins can be assigned based on the intracellular location and the MS shift matching the O-GlcNAc modification.

We therefore sought to develop metabolic labelling of O-GlcNAc in rice seedlings, based on the protocol previously established in our laboratory for labelling of N-linked glycans in *A. thaliana* seedlings (Zhu et al., 2016; Zhu and Chen, 2017). The 14-day-old seedlings of *japonica* rice (*Oryza sativa* L.) cultivar Zhonghua 11 (ZH11) were cultured in Kimura B nutrient solution containing 5 mM GalNAz or GlcNAz for 3 days (Figure 1a). Significant growth inhibition was observed for GlcNAz, yet GalNAz did not induce growth inhibition on rice seedlings (Figure 1b). We therefore selected GalNAz for further experiments. As previously shown by our group as well as others, although plants do not produce GalNAc-containing glycans, GalNAz was able to metabolically label cell surface N-linked glycans (Hoogenboom et al., 2016; Zhu and Chen, 2017), probably by conversion to UDP-GlcNAz through a yet to be uncovered pathway (Furo et al., 2015; Nozaki et al., 2012). The rice seedlings were treated with GalNAz at concentrations ranging from 0.1 to 20 mM for 3 days. The seedlings were then lysed and reacted with alkyne-Cy5 via click chemistry. In-gel fluorescence analysis showed that a variety of proteins were metabolically labelled with azides in a GalNAz-concentration-dependent manner (Figure 1c). The labelling intensity was saturated at 10 mM. The seedlings of Nipponbare (NP) in the *japonica* background were also evaluated for metabolic labelling with GalNAz. As shown by the in-gel fluorescence assay, significant labelling was observed in both ecotypes, with ZH11 showing a slightly higher labelling intensity with 10 mM GalNAz (Figure 1d). Finally, we examined rice gross morphologies under 10 mM GalNAz treatment for 3 days. No abnormal morphological phenotypes were observed in either ZH11 or NP, demonstrating the biocompatibility of the GalNAz labelling (Figure 1e; Figure S2). Taken together, treatment of 14-day-old ZH11 seedlings with 10 mM GalNAz for 3 days was chosen for large-scale identification of O-GlcNAcylated proteins and modification sites in this work.

Profiling of O-GlcNAcylation sites in rice

To identify O-GlcNAcylated proteins and modification sites, lysates of the GalNAz-treated seedlings were reacted with alkyne-AC-biotin, which contains an acid-cleavable dialkoxydiphenylsilane linker (Figure 2a; Figure S3). The click-labelled glycoproteins were captured with streptavidin beads, followed by on-bead trypsin digestion. The beads were then treated with formic acid to release the O-GlcNAcylated peptides, which were analysed by LC–MS/MS. We first compared three fragmentation methods in MS/MS analysis of O-GlcNAcylated peptides, including electron transfer dissociation (ETD; Chalkley et al., 2009), electron transfer/higher-energy collision dissociation (EThcD; Frese et al., 2012) and higher-energy collision product-dependent EThcD (HCD-pd-EThcD; Liu et al., 2021). Among the three fragmentation methods, HCD-pd-EThcD produced the highest number of total peptide spectral matches (PSMs) and HexNAz-containing PSMs (Figure 2b). Furthermore, the highest percentages of HexNAz PSMs and HexNAz peptides were achieved by HCD-pd-EThcD (Figure 2c,d). The best performance of HCD-pd-EThcD on identifying O-GlcNAcylated peptides in rice seedlings was in agreement with our recent evaluation in mammalian cells (Liu et al., 2022).
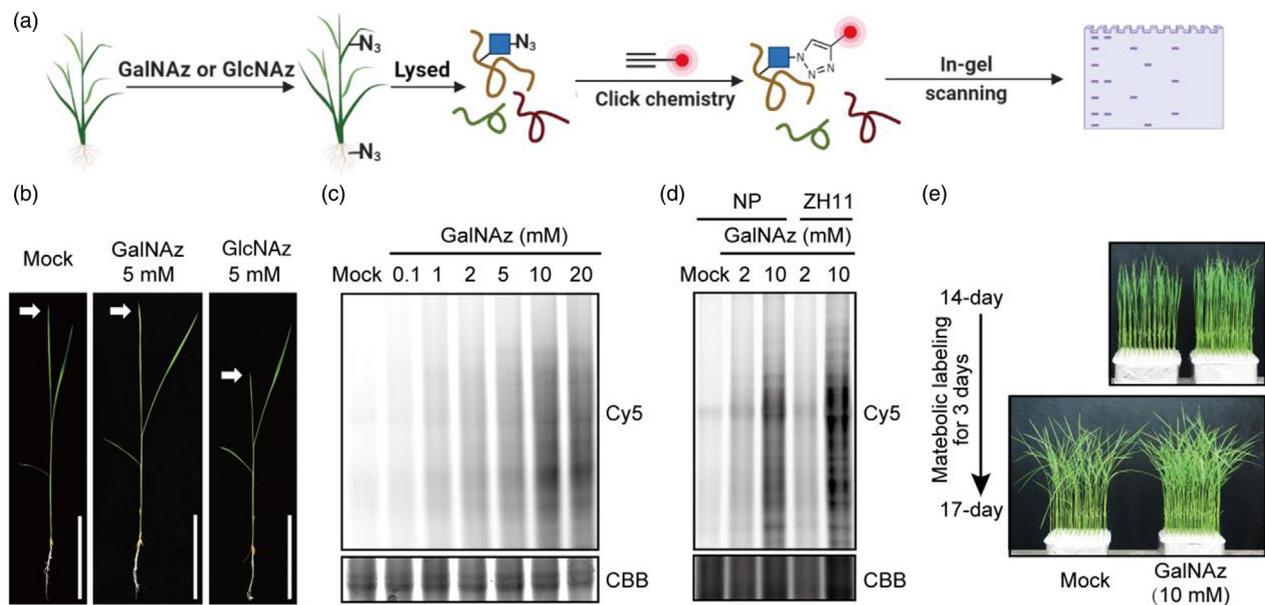
Figure 1 Metabolic glycan labelling with GalNAz and GlcNAz in rice seedlings. (a) MGL enables fluorescence detection of protein glycosylation in rice. Fourteen-day-old rice seedlings are grown with azidosugar (e.g. GalNAz or GlcNAz) for 3 days, which are metabolically incorporated into O-GlcNAc as well as other glycans. The lysates of the azide-incorporated seedlings are reacted with alkyne-fluorophore via click chemistry for in-gel fluorescence scanning analysis. (b) Effects of GlcNAz and GalNAz on the growth of rice seedlings. The 14-day-old ZH11 seedlings were treated with 5 mM GalNAz, GlcNAz, or mock for 3 days, followed by comparison of the seedling height, as indicated by the white arrows. Scale bars, 10 cm. (c) Dependence of labelling on the concentration of GalNAz. The 14-day-old ZH11 seedlings were treated with GalNAz at varied concentrations for 3 days, followed by reaction with alkyne-Cy5 and analysis by in-gel fluorescence scanning. Coomassie brilliant blue (CBB)-stained gel demonstrates comparable loading. (d) MGL with GalNAz in two rice ecotypes. The 14-day-old rice seedlings of NP and ZH11 were treated with GalNAz at 2 or 10 mM for 3 days, followed by reaction with alkyne-Cy5 and analysis by in-gel fluorescence scanning. (e) Gross morphologies of 14-day-old ZH11 seedlings treated with 10 mM GalNAz for 3 days.

Using the ETD, EThcD and HCD-pd-EThcD fragmentation methods, we identified 782, 1059 and 1930 O-GlcNAcylation sites, respectively (Figure S4a,c,e; Data S1–S3). We categorized the identified O-GlcNAcylation sites into two classes according to the PTM score: Class I from 1 to 0.75 and Class II below 0.75. By assigning Class I as unambiguous sites, 566, 538 and 1081 unambiguous sites were identified with three fragmentation methods, respectively (Figure 2e). With HCD-pd-EThcD, 849 sites (44%) fell into Class II, and most of these sites were located in peptides with four or more serine and threonine residues, which probably caused the ambiguous assignment (Figure 2f). Motif analysis revealed that serine was the most enriched residue flanking the putative modification sites (Figure S5). These results indicate that rice O-GlcNAcylation tends to occur on serine/threonine-enriched domains.

By combining the data from three acquisition methods, a total of 1591 unambiguous O-GlcNAcylation sites were identified (Figure S6; Data S1–S4). Motif analysis showed a slight preference for proline at the −2 and −3 residues N-terminal to the modification site (Figure 2g), which are commonly found in animals and plants (Trinidad et al., 2012; Xu et al., 2017, 2019). Of the 1591 unambiguous O-GlcNAcylation sites, serine and threonine accounted for 74.6% and 25.4%, respectively (Figure 2h). Among the O-GlcNAcylated proteins, 377 (53%) contained one identified O-GlcNAcylation site and 159 (22%) had more than three sites (Figure 2i). Interestingly, 18 proteins, such as C2H2 zinc finger proteins and translation initiation factor 4G (OseIF4G), were hyper-O-GlcNAcylated with more than 10 O-GlcNAcylation sites (Figure 2i; Table S1).

Analysis of O-GlcNAcylated proteins in rice

Based on the unambiguous sites and ambiguous sites from three acquisition methods, 811 O-GlcNAcylated proteins were identified (Figure 3a; Figure S4b,d,f; Data S4). More than half of the identified O-GlcNAcylated proteins are localized in the nucleus (Figure 3b; Data S4), which is similar to O-GlcNAcylated proteins in animals and other plants (Song et al., 2019; Xu et al., 2017, 2019). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that O-GlcNAcylated proteins are involved in multiple biological processes such as RNA degradation, mRNA surveillance, spliceosome, plant hormone signal transduction and RNA transport (Figure 3c; Figure S7a). Domain analysis showed that O-GlcNAcylated proteins contained many conserved domains, including DNA binding, RNA binding and protein–protein interaction (Figure S7b). Gene Ontology (GO) analysis with the STRING database (https://string-db.org) revealed that O-GlcNAcylated proteins were enriched in the gene expression pathways such as DNA binding, RNA binding and transcription (Figure 3c,d), which were similarly observed in *A. thaliana* (Xu et al., 2017). For example, proteins including the chromatin remodelling components, transcription factors such as WRKYs, PCFs, and SPLs were identified to be O-GlcNAcylated (Figure 3e). In addition, several key components of plant hormone pathways, including auxin response factors (ARFs), were O-GlcNAcylated (Figure 3e), consistent with the observation of ARFs in *A. thaliana* (Xu et al., 2017). These findings suggest the conservation of O-GlcNAcylation between rice and *A. thaliana*.
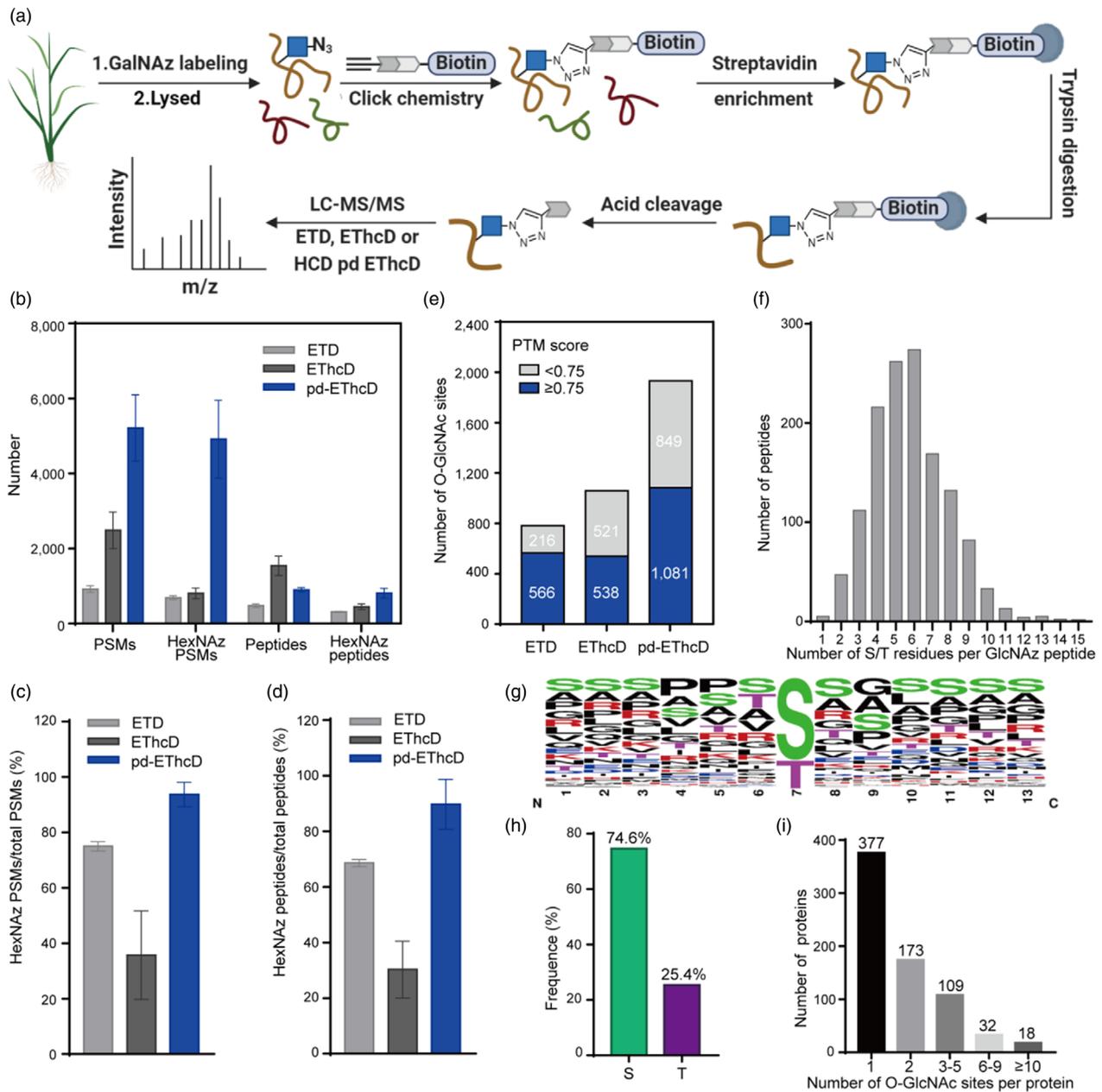
Figure 2 Profiling of O-GlcNAcylation sites. (a) MGL enables proteomic profiling of protein O-GlcNAcylation sites in rice. Rice seedlings are grown with GalNAz for 3 days. The lysates of the azide-incorporated seedlings are reacted with alkyne-AC-biotin, followed by enrichment by streptavidin beads. After on-bead digestion, the O-GlcNAcylated peptides are released by acid cleavage and subjected to LC−MS/MS analysis. (b) Number of PSMs, O-HexNAz PSMs, peptides and O-HexNAz peptides by using the ETD, EThcD and HCD pd EThcD (pd-EThcD) fragmentation methods. Error bars represent means ± SD from three biological repeats. (c and d) The ratio of HexNAz PSMs to total PSMs (c) and the ratio of HexNAz-containing peptides to total identified peptides (d) by using the ETD, EThcD and pd-EThcD fragmentation methods. Error bars represent means ± SD from three biological repeats. (e) Localization probability of O-GlcNAcylation sites. O-GlcNAcylation sites are categorized into two classes according to PTM score: Class I, PTM score ≥0.75; Class II, PTM score <0.75. (f) Frequency distribution of serine (S)/threonine (T) residues per glycopeptide containing ambiguous O-GlcNAcylation sites. (g) Motif analysis of unambiguous O-GlcNAcylation sites by using WebLogo. (h) Proportion of O-GlcNAcylation on serine (S) or threonine (T). (i) Number of O-GlcNAcylation sites per glycoprotein.

Correlation between O-GlcNAcylation and phosphorylation in rice

Both O-GlcNAcylation and phosphorylation occur on the serine and/or threonine residues, and crosstalk between the two PTMs has been studied in animals (Trinidad *et al.*, 2012; Zeidan and Hart, 2010). To explore the relationship between O-GlcNAcylation and phosphorylation in rice, we performed a large-scale identification of phosphorylation sites using rice tissues from the same developmental stage as O-GlcNAcylation profiling. A total of 4585 phosphorylated proteins and 11 657 unambiguous phosphorylation sites were identified (Data S5). By
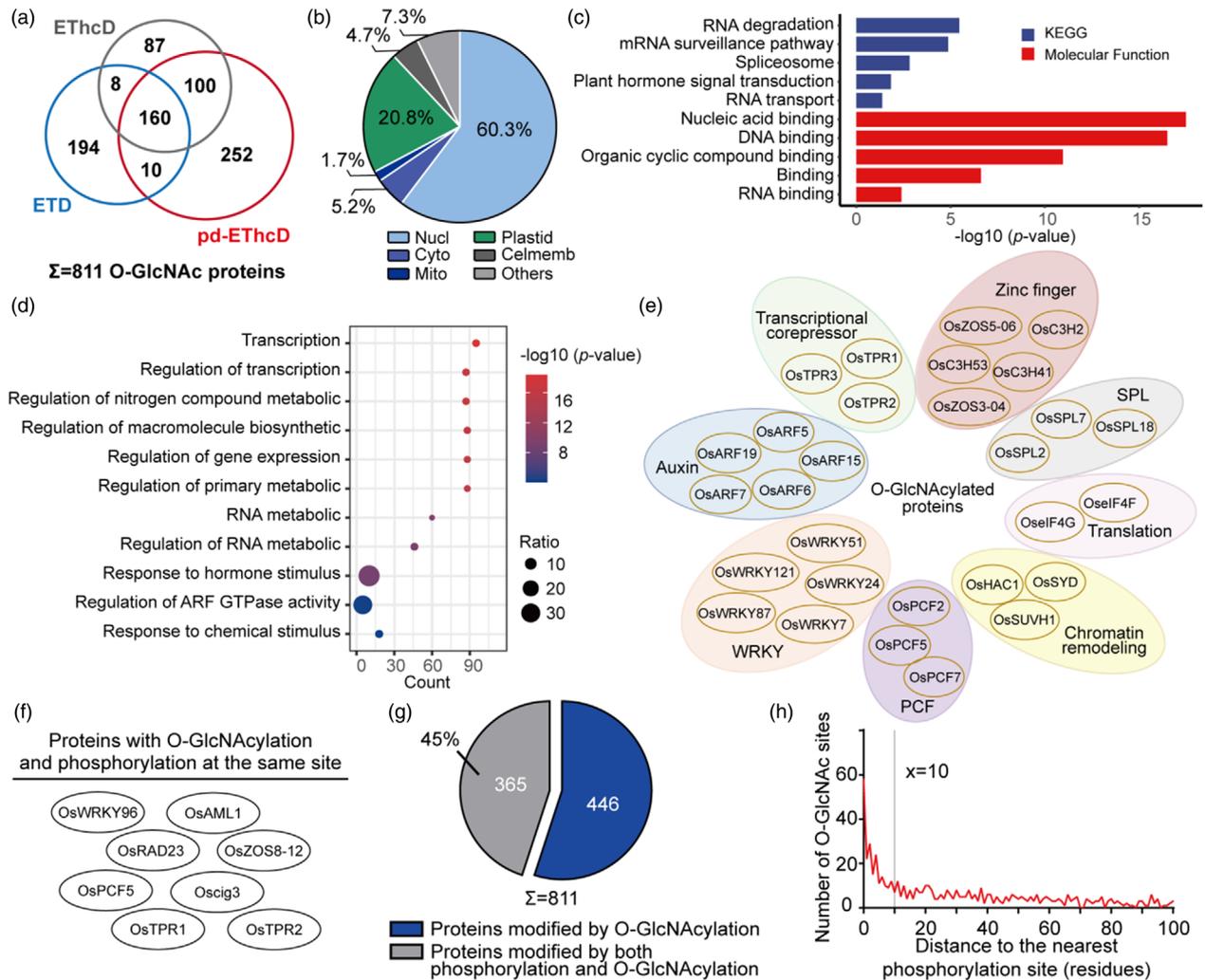
Figure 3 Analysis of O-GlcNAcylated proteins in rice. (a) Venn diagrams showing overlap of O-GlcNAcylated proteins identified with the ETD, EThcD and pd-EThcD fragmentation methods. (b) Pie chart showing the percentages of O-GlcNAcylated proteins in different subcellular locations. Nucl, Cyto, Mito, Plas, Celmemb and Others are abbreviations for nucleus, cytoplasm, mitochondrion, plastid, cell membrane and other subcellular locations, respectively. (c) KEGG pathway and molecular function (MF) analyses of O-GlcNAcylated proteins. (d) Biological processes (BP) analysis of O-GlcNAcylated proteins. (e) Flower plot showing the clusters of O-GlcNAcylated proteins based on their known or predicted functions. (f) Proteins with the O-GlcNAcylation and phosphorylation at the same site. (g) Pie chart showing the percentages of O-GlcNAcylated proteins modified by phosphorylation. (h) Distribution for the nearest distance (amino acid residues) of unambiguous phosphorylation sites to each unambiguous O-GlcNAcylation site. Of 917 unambiguous O-GlcNAcylation sites, 211 sites posited proximity to phosphorylation site within 10 amino acid residues.

comparing O-GlcNAcylation and phosphorylation, 58 sites distributed on 44 proteins were both O-GlcNAcylated and phosphorylated (Figure 3f; Data S6). The 44 proteins included a list of transcription factors such as OsPCF5 and OsWRKY96, suggesting that the crosstalk between O-GlcNAcylation and phosphorylation was involved in transcription. Of the 811 identified O-GlcNAcylated proteins, 45% were also phosphorylated (Figure 3g; Data S6). Gene Ontology and KEGG analyses on the proteins with both O-GlcNAcylation and phosphorylation showed enrichment for RNA degradation, spliceosome and transcription (Figure S8a).

Since O-GlcNAcylation and phosphorylation that simultaneously occur within 10 residues were proposed to incline to crosstalk (Korkuć and Walther, 2017), we calculated the distance of the nearest phosphorylation site to each O-GlcNAc modification. For a total of 211 (21%) O-GlcNAcylation sites, there was a

phosphorylation site within the 10 amino acid residues to that O-GlcNAcylation site (Figure 3h; Figure S8b). In addition, crosstalk may also occur via spatial proximity at the levels of secondary and tertiary structures.

Identification of O-GlcNAc transferase in rice

With hundreds of O-GlcNAcylated proteins identified in rice seedlings, it is of great interest to identify OsOGT. We used the sequence of *A. thaliana* O-GlcNAc transferase SEC to perform a BLAST search against the rice genome, which led to two putative genes, *LOC_Os02g28830* and *LOC_Os01g68680*, designated as *OsOGT1* and *OsOGT2*, respectively. Protein domain analysis showed that both OsOGT1 and OsOGT2 contain an N-terminal protein–protein interaction domain (tetratricopeptide repeats, TPRs) and a C-terminal OGT catalytic domain, which are highly homologous to OGTs in *A. thaliana* and *Homo sapiens* (Figure 4a).

Owing to the lack of 3D structure of plant OGT, we performed pairwise protein structure alignments of OsOGTs with human OGT (HsOGT; Biasini et al., 2014; Lazarus et al., 2013; Lazarus et al., 2011), and generated the predicted 3D structures of OsOGT1 and OsOGT2 (Figure S9a,b). From the predicted structures, we found that OsOGTs and HsOGT both have the conserved catalytic active sites and asparagine ladder (Figure 4b), which are essential for modifying substrate proteins (Levine et al., 2018). These results suggest that OsOGT1 and OsOGT2 should be rice OGTs with a similar catalytic mechanism with human OGT.

To confirm that OsOGT1 and OsOGT2 are *bona fide* O-GlcNAc transferases, we examined their catalytic activities by expressing OsOGT1 and OsOGT2 in *E. coli*. The auto-O-GlcNAcylation of OsOGTs was analysed by using a chemoenzymatic labelling method based on a mutant galactose transferase (Y289L GalT1), which recognizes terminal GlcNAc and transfers a GalNAz moiety from UDP-GalNAz (Clark et al., 2008). After the chemoenzymatic labelling, the lysates were reacted with alkyne-Cy5 via click chemistry. In-gel fluorescence analysis showed that both OsOGT1 and OsOGT2 were auto-glycosylated in *E. coli* (Figure 4c). We

further identified the O-GlcNAcylation sites of OsOGTs by ETD-based MS and found that T59 of OsOGT1, and S382, T446 and S912 of OsOGT2 were O-GlcNAc-modified (Figure 4d,e; Figure S10). Taken together, these results demonstrate that OsOGT1 and OsOGT2 are rice O-GlcNAc transferases.

Verification of O-GlcNAcylated proteins

Finally, we sought to biochemically validate the identified O-GlcNAcylated proteins by modification with OsOGTs. By using a co-expression system (Han et al., 2015), three representative proteins from the list of identified O-GlcNAcylated proteins, including OsRAD23, OsC3H41 and OsC3H53, were co-expressed with OsOGT1 or OsOGT2 in *E. coli* (Figure 5a). The lysates were labelled by the Y289L GalT1-based chemoenzymatic method, which showed that all the three proteins were O-GlcNAcylated in *E. coli* (Figure 5b; Figure S11a). OsRAD23 and OsC3H41 were purified from the *E. coli* co-expression system and subjected to ETD-based MS analysis. S93 of OsRAD23 and S58 of OsC3H41 were identified to be O-GlcNAcylated, which were in agreement with the proteomic results from the GalNAz-treated rice seedlings (Figure S11b–d; Data S4).
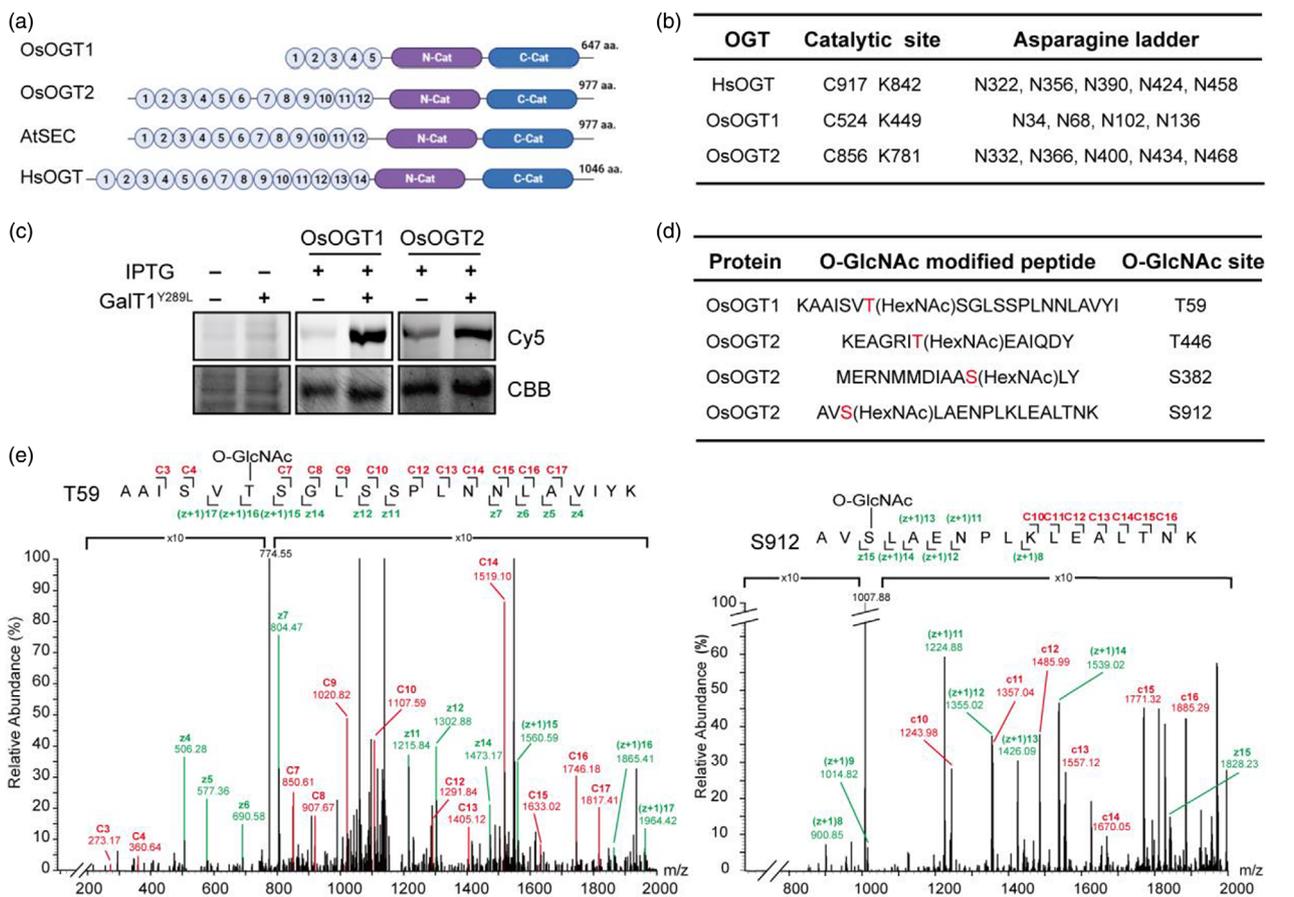

Figure 4 Identification of O-GlcNAc transferase in rice. (a) Protein domain analyses of OsOGT1, OsOGT2, AtSEC and HsOGT using InterPro. Light blue refers to tetratricopeptide repeat (TPR) domains, purple to N-terminal catalytic domains (N-Cat) and dark blue to C-terminal catalytic domains (C-Cat). (b) Conserved catalytic active sites and asparagine ladder of HsOGT and OsOGTs. (c) Auto-glycosylation of OsOGTs. OsOGT1 and OsOGT2 were expressed in *Escherichia coli*, respectively. The bacterial lysates were incubated with or without GalT1^{Y289L} and UDP-GalNAz, followed by reaction with alkyne-Cy5 and analysis by in-gel fluorescence scanning. The CBB-stained gels were used as loading controls. (d) O-GlcNAcylation sites identified from OsOGT1 and OsOGT2 in (c). The modified amino acids are shown in red. (e) MS/MS spectra of the OsOGT1 peptide showing O-GlcNAcylation at T59 (left panel) and the OsOGT2 peptide showing O-GlcNAcylation at S912. The $(z + 1)_{15}$ and $(z + 1)_{16}$ ions indicate O-GlcNAcylation at T59 unambiguously. The $(z + 1)_{14}$ and z_{15} ions indicate O-GlcNAcylation at S912 unambiguously.
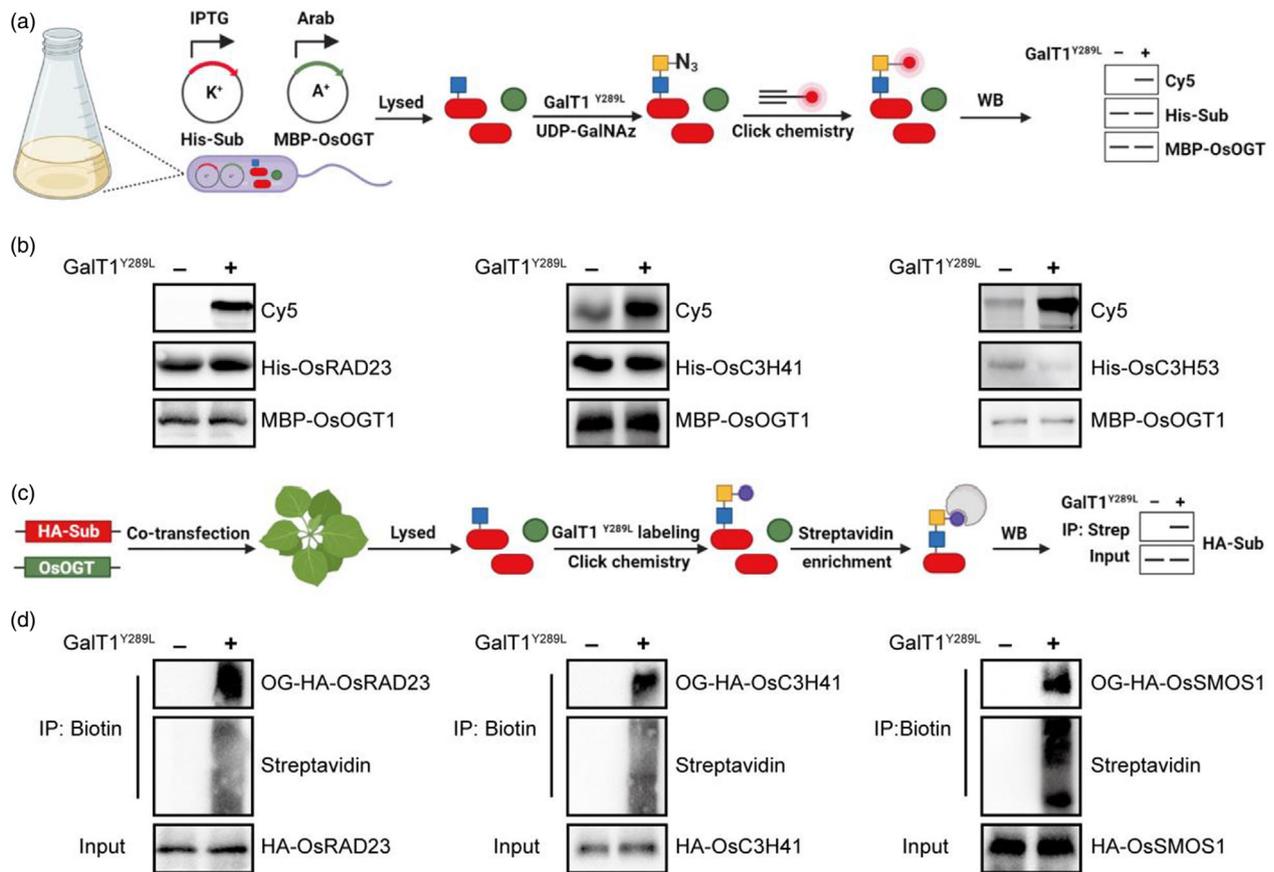
Figure 5 Validation of the O-GlcNAcylated proteins. (a) Schematic of the procedures for O-GlcNAcylated protein validation in *Escherichia coli*. His-tagged OsOGT substrate (His-Sub) and MBP-OsOGT1/2 are co-expressed in *E. coli*. The bacterial lysates are incubated with GalT1^{Y289L} and UDP-GalNAz, followed by reaction with alkyne-Cy5 and analysis by in-gel fluorescence scanning. (b) In-gel fluorescence scanning showing His-OsRAD23, His-OsC3H41 and His-OsC3H53 which were co-expressed with MBP-OsOGT1 in *E. coli*, followed by labelling of O-GlcNAc with Cy5. Immunoblots showing the expression of OsOGT1 and the substrate proteins. (c) Schematic of procedures for O-GlcNAcylated protein validation in tobacco. HA-tagged OsOGT substrate (HA-Sub) and OsOGT1 are co-expressed in tobacco leaves. The tobacco leaf lysates are incubated with GalT1^{Y289L} and UDP-GalNAz, followed by reaction with alkyne-PEG4-biotin. After capture with streptavidin beads, the O-GlcNAcylated proteins are detected by immunoblotting using an anti-HA antibody. (d) Immunoblots showing HA-OsRAD23, HA-OsC3H41 and HA-OsSMOS1 which were co-expressed with OsOGT1 in tobacco leaves, followed by labelling of O-GlcNAc with biotin and capture with streptavidin beads. Immunoblots of the inputs showing comparable loading.

In addition to verifying rice O-GlcNAcylated proteins in the prokaryotic system, we also validated them in the plant system. Using a similar strategy, HA-OsRAD23, HA-OsC3H41 and HA-OsSMOS1 were co-expressed with OsOGT1 in *N. benthamiana* leaves (Figure 5c). The lysates were treated with Y289L GalT1 and UDP-GalNAz, followed by click reaction with alkyne-biotin and capture with streptavidin beads. The streptavidin blot analysis showed that a variety of proteins in *N. benthamiana* were O-GlcNAcylated (Figure 5d). Furthermore, immunoblot using an antibody against the HA tag showed that all three proteins were O-GlcNAcylated in *N. benthamiana* (Figure 5d, upper panels). Collectively, these results demonstrate that our proteomic dataset provides a landscape of O-GlcNAc modification in rice and further support that OsOGT1 and OsOGT2 are OGTs in rice.

In summary, we present the large-scale identification of O-GlcNAcylated proteins in a site-specific manner in rice, which provides a valuable resource for studying O-GlcNAc biology in rice. Furthermore, this work extends the MGL method to rice for the first time. Compared to the O-GlcNAc proteomics studies in *A. thaliana* and wheat using the LWAC enrichment method, MGL in rice resulted in much higher numbers of identified O-GlcNAcylated proteins and modification sites, probably providing better coverage of the O-GlcNAc proteome in plants. Importantly, we discovered two OGTs in rice and biochemically confirmed their enzymatic activity as O-GlcNAc transferases. Considering that SEC might be the only OGT in *A. thaliana* and SPY is an O-fucose transferase, the fact that rice has two OGTs suggests that O-GlcNAc biology has some species-dependent diversity, which is an interesting direction for future studies.

Experimental procedures

Plant materials

Rice (*Oryza sativa* L.) ssp. *japonica* Zhonghua 11 (ZH11) and Nipponbare (NP) were grown in the greenhouse of the Institute of Genetics and Developmental Biology under 16 h light and 8 h dark at 28 °C. Fourteen-day-old seedlings were used for metabolic glycan labelling.

Plasmid construction

To construct prokaryotic expression His-OsOGT plasmids, the coding sequences of *OsOGT1* (*LOC_Os02g28830*) and *OsOGT2* (*LOC_Os01g68680*) were amplified and cloned into the pET28a (+) vector. To construct prokaryotic expression MBP-OsOGT plasmids, the coding sequences of *OsOGT1* and *OsOGT2* were amplified and cloned into the pUCBAD-MBP vector. To construct prokaryotic expression His-Sub plasmids, the coding sequences of *OsRAD23* (*LOC_Os06g15360*), *OsC3H41* (*LOC_Os06g21390*) and *OsC3H53* (*LOC_Os07g48410*) were amplified and cloned into the pET-28a (+) vector. To construct eukaryotic expression 35 S:: OsOGT1 and 35 S::HA-Sub plasmids, the coding sequence of *OsOGT1* was amplified and cloned into the pCAMBIA1300 vector, and the coding sequences of *OsRAD23*, *OsC3H41* and *OsSMOS1* (*LOC_Os05g32270*) were amplified and cloned into the pCAMBIA1300 vector with a 3 × HA tag in the N-terminus. All primers are listed in Table S2.

Chemicals and reagents

DADPS biotin alkyne (catalogue no. 1331), 2-[4-{(bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl}-1H-1,2,3-triazol-1-yl] acetic acid (BTTAA) (catalogue no. 1236), alkyne-PEG4-biotin (catalogue no. TA105) and alkyne-Cy5 (catalogue no. TA116) were purchased from Click Chemistry Tools, GalNAz (catalogue no. 869186-83-4) from Jinan Samuel Pharmaceutical Co., Ltd, streptavidin-conjugated agarose beads (catalogue no. 20350) from Thermo Fisher. Dithiothreitol (DTT) (catalogue no. 43815), ammonium bicarbonate (ABC) (catalogue no. 09830), iodoacetamide (IAA) (catalogue no. I1149), urea (catalogue no. U5128) and formic acid (FA; catalogue no. 06473) were purchased from Sigma Aldrich, and sequencing-grade modified trypsin (catalogue no. V5111) and trypsin resuspension buffer (catalogue no. V542A) were purchased from Promega. All organic solvents were analytical grade or better. Antibodies included anti-HA (CST, #2367, 1 : 3000), anti-His (MBL, D291-3, 1 : 10 000), anti-MBP, (Proteintech, 15 089-1-AP, 1 : 10 000), anti-streptavidin-HRP (Beyotime, A0303, 1 : 5000), anti-mouse IgG-HRP (CST, #7076, 1 : 10 000) and anti-rabbit IgG-HRP (CST, #7074, 1 : 10 000).

Metabolic labelling of living plants

Fourteen-day-old seedlings were grown in Kimura B nutrient solution with indicated unnatural sugars at various concentrations for 3 days and then harvested in liquid nitrogen. Lysis buffer containing 50 mM sodium phosphate buffer, pH 7.0, 150 mM NaCl, 10% glycerol, 0.1% NP-40 and complete protease inhibitor cocktail was added to the frozen samples, followed by homogenization and centrifugation. The supernatants were filtered through 0.2-μm filter and quantified by BCA Protein Assay Kit (Pierce, catalogue no. 23225).

Chemoenzymatic labelling of O-GlcNAcylated proteins

Cell lysates were centrifuged at 18 000 *g* for 20 min at 4 °C, and then, the supernatants were precipitated with methanol chloroform. After centrifugation, the pellets were re-dissolved in the buffer containing 20 mM HEPES (pH 7.9), 1% SDS (wt/vol) to a final concentration of 2–5 mg protein per 1 mL. Forty microlitres lysates were mixed with 49 μL Milli-Q water, 80 μL labelling buffer [125 mM NaCl, 5% NP-40 (vol/vol), 50 mM HEPES, pH 7.9], 11 μL 100 mM $MnCl_2$, 10 μL 500 μM UDP-GalNAz and 7 μL GalT1^{Y289L}. The negative control was performed without GalT1^{Y289L}. The mixture was gently rotated for 20 h at 4 °C.

The reaction was precipitated with methanol chloroform (3 : 1, vol/vol) and then re-dissolved in 50 μL PBS buffer containing 1% SDS (wt/vol) for click reaction.

Click chemistry

The unnatural sugars or chemoenzymatically labelled samples were precipitated with methanol chloroform. The protein pellets were re-dissolved in 50 μL PBS buffer containing 1% SDS (wt/vol); then, 50 μM $CuSO_4$-BTTAA premixed complex ($CuSO_4$-BTTAA, molar ratio 1 : 2), 100 μM alkyne-probes and 12.5 mM fresh sodium ascorbate were added for click reaction at 25 °C for 2 h. The reaction product was used for in-gel fluorescence scanning by Typhoon FLA 9500 Fluorescence Imager (GE Healthcare, Life Sciences) or precipitated with methanol chloroform to remove click reagents for further O-GlcNAcylation site identification.

Enrichment of O-GlcNAcylated proteins

The precipitated pellets after click reaction were re-dissolved in 1 mL PBS buffer containing 1.2% SDS and then diluted SDS with PBS buffer to 0.2%. Two hundred microlitres pre-washed streptavidin-conjugated agarose beads were added to protein solution and incubated at room temperature for 4 h. The beads were washed sequentially with PBS (pH 7.4) and Milli-Q water for five times to remove non-specifically bound proteins, then re-suspended in 2 × SDS loading buffer and analysed by western blotting. For MS/MS analysis, the washed beads were re-suspended in 6 M urea.

Preparation of protein samples for mass spectrometry

Ten-milligram GalNAz labelled protein lysates were used for click reaction with DADPS biotin alkyne and followed by enrichment as mentioned above. After enrichment, the streptavidin-conjugated beads were re-suspended in 500 μL PBS buffer containing 6 M urea and 10 mM DTT, and incubated at 37 °C for more than 30 min. Iodoacetamide (25 μL, 400 mM) was then added and incubated at 37 °C for 45 min in the dark. The beads were washed with PBS buffer followed by centrifugation and re-suspended in 200 μL PBS buffer containing 2 M urea, 1 mM $CaCl_2$ and 10 μg trypsin. Trypsin digestion was performed on a rotary shaker at 37 °C for more than 18 h. After digestion, the beads were washed sequentially with PBS buffer (pH 7.4) and Milli-Q water for five times. To release modified peptides, the beads were re-washed with 200 μL 2% (v/v) formic acid (FA)/water for 2 h at room temperature. The eluents were collected, and beads were washed with 50% (v/v) acetonitrile (ACN)/water containing 1% FA (200 μL) and 200 μL Milli-Q water. All the eluents were combined and dried in a vacuum centrifuge and then stored at −30 °C.

High-pH stage tips for fractionation

To acquire high-quality ETD and EThcD data, samples were performed with high-pH stage tips for fractionation (Rappsilber *et al.*, 2007). Briefly, C18 membrane (3 M) was inserted into 10-μL tips (Axygen) followed by adding 200 μg Durashell C18 beads (Agela), which were suspended in ACN. The beads were washed by 80% and 50% ACN in 10 mM pH 10.0 ammonium bicarbonate (ABC) twice and subsequently balanced with 10 mM ABC (pH 10.0) twice. The dried peptides were dissolved in 100 μL of 10 mM ABC (pH 10.0) and loaded onto the prepared tips more than five times. The tips were washed twice with 50 μL 10 mM ABC (pH 10.0), and then, the peptides were eluted with 6%, 9%,

12%, 15%, 18%, 21%, 25%, 30%, 35%, 40% and 80% ACN in 10 mM ABC (pH 10.0). A total of 11 fractions were concatenated into six fractions by combining fraction 1 and 7; 2 and 8; 3 and 9; 4 and 10; 5 and 11. Finally, the six fractions and the flow through were dried in a vacuum centrifuge and stored at −30 °C.

LC–MS/MS analysis

The dried peptides were reconstituted in 0.2% FA and separated by applying a loading column (100 μm × 2 cm), followed by a C18 separation capillary column (75 μm × 15 cm) packed in-house with Luna 3 μm C18 bulk packing material (Phenomenex, Beijing). For ETD data, the mobile phases (A: water with 0.1% FA, B: 100% ACN with 0.1% FA) were driven and controlled by an EASY-nLC 1000 system (Thermo Fisher Scientific). The liquid chromatography gradient was held at 2% B for 1 min, followed by an increase from 2% to 7% B in 1 min, an increase from 7% to 35% B in 120 min and an increase from 35% to 75% B in 4 min. For EThcD data, the mobile phases (A: water with 0.1% FA, B: 80% ACN with 0.1% FA) were driven and controlled by a Dionex Ultimate™ 3000 RPLCnano system (Thermo Fisher Scientific). The liquid chromatography gradient was held at 2% B for 8 min, followed by an increase from 2% to 44% B in 70 min and an increase from 44% to 99% B in 5 min. For HCD-pd-EThcD data, the mobile phases (A: water with 0.1% FA, B: 80% ACN with 0.1% FA) were driven and controlled by a Dionex Ultimate™ 3000 RPLCnano system (Thermo Fisher Scientific). The liquid chromatography gradient was held at 1% B for 8 min, followed by an increase from 1% to 7% B from 8 to 9 min, an increase from 7% to 35% B from 9 to 311 min, an increase from 35% to 44% B from 311 to 353 min and an increase from 44% to 99% B from 353 to 356 min. For ETD data, the mass spectrometry data were acquired in data-dependent acquisition mode with a full MS scan (300–1700 m/z) in FT mode with a resolution of 60 000 followed by ETD MS/MS scans on the 10 most abundant ions with multiple charges in the initial MS scan. Automatic gain control (AGC) targets were 1e6 ions for Orbitrap scans and 5e4 for MS/MS scans. For dynamic exclusion, the following parameters were used: isolation window, 2 m/z; repeat count, 1; repeat duration, 25 s; exclusion duration, 25 s. The ETD activation time was 150 ms. Charge state-dependent time and supplemental activation for ETD were enabled. For EThcD and HCD-pd-EThcD data, the samples were analysed by Orbitrap Fusion™ LUMOS™ Tribrid™ Mass Spectrometer, the precursors were ionized using an EASY-Spray™ ionization source (Thermo Fisher Scientific) held at +2.0 kV compared with ground, and the inlet capillary temperature was held at 320 °C. Survey scan of peptide precursors was collected in the Orbitrap from 350 to 1800 m/z (EThcD) and 350 to 2000 m/z (HCD-pd-EThcD) with an AGC target of 400 000 (EThcD) and 500 000 (HCD-pd-EThcD), a maximum injection time of 50 ms, RF lens at 30% (EThcD) and 60% (HCD-pd-EThcD), a resolution of 120 000 at 200 m/z. Monoisotopic precursor selection was enabled for peptide isotopic distributions, precursors of $z = 2–8$ were selected for data-dependent MS/MS scans for 3 s of cycle time, and dynamic exclusion was set to 15 s with a ±10 ppm window set around the precursor monoisotope. For HCD-pd-EThcD data, the 'scout HCD' scans used an automated scan range determination. The first mass of 120 Th, a normalized collision energy (nce) of 30 ± 10, an AGC target value of 50 000, a maximum injection time of 54 ms and a resolution of 30 000 at 200 m/z were selected. If fragment ions with m/z 329.1461 were detected with a mass error within 10 ppm, product-dependent EThcD scans

would be triggered. In EThcD scans, The first mass of 120 Th, an SA collision energy of 35%, an AGC target value of 10 000, a maximum injection time of 200 ms and a resolution of 30 000 at 200 m/z were selected.

Data analysis

The ETD, EThcD and HCD-pd-EThcD raw data were processed by the MaxQuant software suite (version 1.6.2.10). Rice proteome database was downloaded from MUS on 4 July, 2017, and used for full MS and tandem MS/MS search by the Andromeda Search Engine in the MaxQuant framework. The parameters of database search were as described previously (Qin et al., 2018) with a minimal modification. Briefly, the precursor mass tolerance was 20 ppm for first search and 4.5 ppm for main search, with a fragment mass tolerance of 0.5 Da. Enzyme digestion was set to trypsin with two missed cleavages and the minimum peptide length of seven amino acids. A false discovery rate (FDR) of 0.01 was required for proteins and peptides. The variable modification was set up as HexNAz tagged adduct (328.1383 Da) assigned to serine, threonine and asparagine, and carbamidomethyl was chosen as a fixed modification. For site identification, peptides with the Andromeda score ≥ 40 and delta score ≥6 were regarded as modified peptides. The modified peptide containing HexNAz moiety on asparagine was regarded as N-HexNAz and excluded. The spectra with a PTM score ≥0.75 were considered as the unambiguous sites with high confidence, and those with PTM score <0.75 were considered as ambiguous sites. The proteins containing either unambiguous sites or ambiguous sites were regarded as O-GlcNAcylated proteins.

Phosphopeptides enrichment and LC–MS/MS

The powder of 17-day-old rice seedlings was lysed in a buffer containing 4% SDS and 100 mM Tris–HCl (pH 8.5). After incubation at 95 °C for 10 min, the lysates were centrifuged for 10 min at 12 000 g to remove the insoluble debris. Proteins (~1200 μg) were reduced and alkylated by TCEP and CAA, and digested with trypsin (1 : 50 w/w) at 37 °C overnight. After digestion, the phosphopeptides were enriched using titanium dioxide beads (TiO$_2$; GL Sciences, 5010-21315). The phospho-peptides were analysed by an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer coupled online to an Easy-nLC 1000 in the data-dependent mode. Precursor ions were measured in the Orbitrap analyser at 240 000 resolution. The 20 most intense ions from each MS scan were isolated and fragmented by high-energy collisional dissociation. The database search was performed as mentioned above. Serine, threonine and tyrosine phosphorylation, protein N-terminal acetylation and methionine oxidation were included in the search as the variable modifications. Cysteine carbamidomethylation was set as stable modifications. For site identification, peptides with the Andromeda score ≥ 40 and delta score ≥6 were regarded as modified peptides. The spectra with PTM score ≥0.75 were considered as the unambiguous sites with high confidence, and those with PTM score <0.75 were considered as ambiguous sites. The protein containing either unambiguous site or ambiguous site was regarded as phosphorylated protein.

Bioinformatics

For subcellular location analysis, protein sequences were downloaded from the Rice Genome Annotation Project (https://rice.plantbiology.msu.edu/index.shtml) and predicted by Plant-mSubP (https://bioinfo.usu.edu/Plant-mSubP/). The unambiguous sites

were used for motif analysis, and sequence logos were created by Weblogo (https://weblogo.berkeley.edu/logo.cgi). The KEGG pathway analysis was performed by STRING (https://string-db.org/). Gene Ontology enrichment was performed using agriGO (https://bioinfo.cau.edu.cn/agriGO/index.php).

Identification of O-GlcNAcylation sites of OsOGTs

Rice OGT coding sequences were cloned into the pET-28a (+) vector and then transformed into *E. coli* BL21 (DE3). *Escherichia coli* cells were grown in LB medium at 37 °C to $OD_{600} = 0.6$ and then, isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression overnight at 18 °C. The *E. coli* cells were harvested and lysed in PBS buffer containing complete protease inhibitor cocktail, followed by sonication. The lysates were separated into two parts. One part was subjected to chemoenzymatic labelling to detect O-GlcNAc modification by in-gel fluorescence, and the other part was resolved by 8% SDS-PAGE for MS to identify O-GlcNAcylation sites. Briefly, the gels were stained with Coomassie brilliant blue (CBB) for cutting. The gel pieces were dehydrated in ACN, incubated in 50 mM ABC buffer containing 10 mM DTT for 40 min at 56 °C, incubated in 50 mM ABC buffer containing 55 mM IAA for 1 h in the dark at room temperature and dehydrated again sequentially. The gel pieces were in-gel digested with 5–10 ng/µL trypsin or chymotrypsin in 50 mM ABC overnight at 37 °C. The peptides were extracted twice with 5% FA over 50% ACN and vacuum-centrifuged to dryness. All samples were re-dissolved in 0.1% FA water prior to LC–MS/MS analysis.

Verification of rice O-GlcNAcylated proteins in *E. coli*

To verify O-GlcNAcylated proteins identified in this study, OsOGTs and O-GlcNAcylated proteins were co-expressed in an *E. coli* dual plasmid system. Briefly, coding sequences of *OsOGTs* were cloned into the pUCBAD-MBP vector with ampicillin resistance and coding sequences of O-GlcNAcylated proteins were cloned into the pET28a (+) vector with kanamycin resistance. The pUCBAD-MBP plasmid containing the *E. coli* arabinose operon promoter to express recombinant MBP-OsOGT and the pET28a (+) plasmid in which the expression of recombinant His-Sub was controlled by the T7 promoter and lac operator were cotransformed into *E. coli* BL21 (DE3) cells. To express OsOGTs and O-GlcNAcylated proteins, the cotransformed *E. coli* cells were grown in 1 mL of LB medium at 37 °C to $OD_{600} = 0.6$ and then arabinose and IPTG were added to induce protein expression overnight at 18 °C with the final concentrations of 0.5% (w/v) and 1 mM, respectively. Total proteins were extracted from *E. coli* with PBS buffer, precipitated with methanol chloroform and then re-dissolved in 200 µL 20 mM HEPES (pH 7.9) containing 1% SDS (w/v). Forty microlitres lysate was used for chemoenzymatic labelling as described previously. For O-GlcNAcylation site identification, the lysates were resolved by 10% SDS-PAGE. The gels were stained with CBB for cutting. In-gel digestion was performed as described previously.

Verification of rice O-GlcNAcylated proteins in tobacco

To validate rice O-GlcNAcylated proteins, the coding sequence of *OsOGT1* was cloned into the binary vector pCAMBIA1300 while the coding sequences of O-GlcNAcylated proteins were cloned into pCAMBIA1300 with a 3 × HA tag in the N-terminus. The pCAMBIA1300-OsOGT1 and pCAMBIA1300-HA-Sub vectors were introduced into *Agrobacterium tumefaciens* EHA105. *Agrobacterium tumefaciens* cells were diluted with infiltration buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$, 200 µM acetosyringone) to $OD_{600} = 0.7$ and then injected into tobacco (*Nicotiana benthamiana*) leaves, which were grown under a 16-h light/8-h dark photoperiod for 48 h at 23 °C. Injected leaves were harvested in liquid nitrogen and ground into powder. Three-gram powder was lysed in lysis buffer containing 50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 10% glycerol, 0.1% NP-40 and complete protease inhibitor cocktail. The lysates were precipitated with methanol chloroform and re-dissolved in 1 mL 20 mM HEPES (pH 7.9) containing 1% SDS (w/v). Then, 400 µL lysates were used for chemoenzymatic labelling and followed by a click reaction for purification of biotinylated proteins.

Conflict of interest

The authors declare no competing financial interests.

Author contributions

X.L., C.L., X.C. and J.L. conceived and designed the study. X.L., C.L., Q.S., L.B., B.C., K.Q., X.L., B.M., B.W and W.Z. performed the experiments. X.L. and C.L. performed the data analysis. X.L., C.L., X.C. and J.L. wrote the manuscript. All authors approved the final manuscript.

Data availability statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD036527.

References

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258.

Bond, M.R. and Hanover, J.A. (2013) O-GlcNAc cycling: a link between metabolism and chronic disease. *Annu. Rev. Nutr.* **33**, 205–229.

Bond, M.R. and Hanover, J.A. (2015) A little sugar goes a long way: the cell biology of O-GlcNAc. *J. Cell Biol.* **208**, 869–880.

Chalkley, R.J., Thalhammer, A., Schoepfer, R. and Burlingame, A.L. (2009) Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. USA* **106**, 8894–8899.

Chen, R., Deng, Y., Ding, Y., Guo, J., Qiu, J., Wang, B., Wang, C. *et al.* (2022) Rice functional genomics: decades' efforts and roads ahead. *Sci. China Life Sci.* **65**, 33–92.

Clark, P.M., Dweck, J.F., Mason, D.E., Hart, C.R., Buck, S.B., Peters, E.C., Agnew, B.J. *et al.* (2008) Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J. Am. Chem. Soc.* **130**, 11576–11577.

El Sayed, M.A.A., Kheir, A.M.S., Hussein, F.A., Ali, E.F., Selim, M.E., Majrashi, A. and El Shamey, E.A.Z. (2021) Developing new lines of *Japonica* rice for higher quality and yield under arid conditions. *PeerJ* **9**, e11592.

Frese, C.K., Altelaar, A.F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J. and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668–9673.

Fukagawa, N.K. and Ziska, L.H. (2019) Rice: importance for global nutrition. *J. Nutr. Sci. Vitaminol. (Tokyo)* **65**, S2–s3.

Furo, K., Nozaki, M., Murashige, H. and Sato, Y. (2015) Identification of an *N*-acetylglucosamine kinase essential for UDP-*N*-acetylglucosamine salvage synthesis in *Arabidopsis. FEBS Lett.* **589**, 3258–3262.

Gao, Y., Wells, L., Comer, F.I., Parker, G.J. and Hart, G.W. (2001) Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-*N*-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838–9845.

Goff, S.A. (1999) Rice as a model for cereal genomics. *Curr. Opin. Plant Biol.* **2**, 86–89.

Haltiwanger, R.S., Holt, G.D. and Hart, G.W. (1990) 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.

Han, C., Shan, H., Bi, C., Zhang, X., Qi, J., Zhang, B., Gu, Y. *et al.* (2015) A highly effective and adjustable dual plasmid system for O-GlcNAcylated recombinant protein production in *E. coli. J. Biochem.* **157**, 477–484.

Hartweck, L.M., Scott, C.L. and Olszewski, N.E. (2002) Two O-linked *N*-acetylglucosamine transferase genes of *Arabidopsis thaliana* L. Heynh. have overlapping functions necessary for gamete and seed development. *Genetics* **161**, 1279–1291.

Hoogenboom, J., Berghuis, N., Cramer, D., Geurts, R., Zuilhof, H. and Wennekes, T. (2016) Direct imaging of glycans in *Arabidopsis* roots via click labeling of metabolically incorporated azido-monosaccharides. *BMC Plant Biol.* **16**, 220.

Korkuć, P. and Walther, D. (2017) Towards understanding the crosstalk between protein post-translational modifications: Homo- and heterotypic PTM pair distances on protein surfaces are not random. *Proteins* **85**, 78–92.

Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P. and Walker, S. (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567.

Lazarus, M.B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W.F., Vocadlo, D.J. *et al.* (2013) HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **342**, 1235–1239.

Levine, Z.G., Fan, C., Melicher, M.S., Orman, M., Benjamin, T. and Walker, S. (2018) O-GlcNAc transferase recognizes protein substrates using an asparagine ladder in the tetratricopeptide repeat (TPR) superhelix. *J. Am. Chem. Soc.* **140**, 3510–3513.

Li, B. and Kohler, J.J. (2014) Glycosylation of the nuclear pore. *Traffic* **15**, 347–361.

Liu, J., Hao, Y., He, Y., Li, X., Sun, D.E., Zhang, Y., Yang, P.Y. *et al.* (2021) Quantitative and site-specific chemoproteomic profiling of protein O-GlcNAcylation in the cell cycle. *ACS Chem. Biol.* **16**, 1917–1923.

Liu, J., Hao, Y., Wang, C., Jin, Y., Yang, Y., Gu, J. and Chen, X. (2022) An optimized isotopic photocleavable tagging strategy for site-specific and quantitative profiling of protein O-GlcNAcylation in colorectal cancer metastasis. *ACS Chem. Biol.* **17**, 513–520.

Lubas, W.A., Frank, D.W., Krause, M. and Hanover, J.A. (1997) O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9316–9324.

Ma, J., Li, Y., Hou, C. and Wu, C. (2021) O-GlcNAcAtlas: A database of experimentally identified O-GlcNAc sites and proteins. *Glycobiology* **31**, 719–723.

Nozaki, M., Sugiyama, M., Duan, J., Uematsu, H., Genda, T. and Sato, Y. (2012) A missense mutation in the glucosamine-6-phosphate *N*-acetyltransferase-encoding gene causes temperature-dependent growth defects and ectopic lignin deposition in *Arabidopsis. Plant Cell* **24**, 3366–3379.

Olszewski, N.E., West, C.M., Sassi, S.O. and Hartweck, L.M. (2010) O-GlcNAc protein modification in plants: evolution and function. *Biochim. Biophys. Acta* **1800**, 49–56.

Pucci, M., Malagolini, N. and Dall'Olio, F. (2021) Glycobiology of the epithelial to mesenchymal transition. *Biomedicine* **9**, 770.

Qin, W., Lv, P., Fan, X., Quan, B., Zhu, Y., Qin, K., Chen, Y. *et al.* (2017) Quantitative time-resolved chemoproteomics reveals that stable O-GlcNAc regulates box C/D snoRNP biogenesis. *Proc. Natl. Acad. Sci. USA* **114**, E6749–E6758.

Qin, K., Zhu, Y., Qin, W., Gao, J., Shao, X., Wang, Y.L., Zhou, W. *et al.* (2018) Quantitative profiling of protein O-GlcNAcylation sites by an isotope-tagged cleavable linker. *ACS Chem. Biol.* **13**, 1983–1989.

Qin, W., Xie, Z., Wang, J., Ou, G., Wang, C. and Chen, X. (2020) Chemoproteomic profiling of O-GlcNAcylation in *Caenorhabditis elegans. Biochemistry* **59**, 3129–3134.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906.

Salomon, D. and Orth, K. (2013) What pathogens have taught us about posttranslational modifications. *Cell Host Microbe* **14**, 269–279.

Sasaki, A.T., Carracedo, A., Locasale, J.W., Anastasiou, D., Takeuchi, K., Kahoud, E.R., Haviv, S. *et al.* (2011) Ubiquitination of K-Ras enhances activation and facilitates binding to select downstream effectors. *Sci. Signal.* **4**, ra13.

Schjoldager, K.T., Narimatsu, Y., Joshi, H.J. and Clausen, H. (2020) Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749.

Schneider, P. and Asch, F. (2020) Rice production and food security in Asian Mega deltas—A review on characteristics, vulnerabilities and agricultural adaptation options to cope with climate change. *J. Agronomy Crop Sci.* **206**, 491–503.

Slawson, C. and Hart, G.W. (2011) O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer* **11**, 678–684.

Song, H., Ma, J., Bian, Z., Chen, S., Zhu, J., Wang, J., Huang, N. *et al.* (2019) Global profiling of O-GlcNAcylated and/or phosphorylated proteins in hepatoblastoma. *Signal Transduct. Target. Ther.* **4**, 40.

Trinidad, J.C., Barkan, D.T., Gulledge, B.F., Thalhammer, A., Sali, A., Schoepfer, R. and Burlingame, A.L. (2012) Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol. Cell. Proteomics* **11**, 215–229.

Walsh, C.T., Garneau-Tsodikova, S. and Gatto, G.J., Jr. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372.

Wells, L., Vosseller, K. and Hart, G.W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291**, 2376–2378.

Wong, C.H. (2005) Protein glycosylation: new challenges and opportunities. *J. Org. Chem.* **70**, 4219–4225.

Woo, C.M. and Bertozzi, C.R. (2016) Isotope targeted glycoproteomics (IsoTaG) to characterize intact, metabolically labeled glycopeptides from complex proteomes. *Curr. Protoc. Chem. Biol.* **8**, 59–82.

Woo, C.M., Lund, P.J., Huang, A.C., Davis, M.M., Bertozzi, C.R. and Pitteri, S.J. (2018) 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.

Xiao, J., Xu, S., Li, C., Xu, Y., Xing, L., Niu, Y., Huan, Q. *et al.* (2014) O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits *TaVRN1* mRNA accumulation during vernalization in winter wheat. *Nat. Commun.* **5**, 4572.

Xu, S.L., Chalkley, R.J., Maynard, J.C., Wang, W., Ni, W., Jiang, X., Shin, K. *et al.* (2017) Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in *Arabidopsis. Proc. Natl. Acad. Sci. USA* **114**, E1536–E1543.

Xu, S., Xiao, J., Yin, F., Guo, X., Xing, L., Xu, Y. and Chong, K. (2019) The protein modifications of O-GlcNAcylation and phosphorylation mediate vernalization response for flowering in winter wheat. *Plant Physiol.* **180**, 1436–1449.

Yang, X. and Qian, K. (2017) Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **18**, 452–465.

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., Chen, M. *et al.* (2021) A route to de novo domestication of wild allotetraploid rice. *Cell* **184**, 1156–1170.e1114.

Yuzwa, S.A. and Vocadlo, D.J. (2014) O-GlcNAc and neurodegeneration: biochemical mechanisms and potential roles in Alzheimer's disease and beyond. *Chem. Soc. Rev.* **43**, 6839–6858.

Zeidan, Q. and Hart, G.W. (2010) The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J. Cell Sci.* **123**, 13–22.

Zentella, R., Hu, J., Hsieh, W.P., Matsumoto, P.A., Dawdy, A., Barnhill, B., Oldenhof, H. *et al.* (2016) O-GlcNAcylation of master growth repressor DELLA by SECRET AGENT modulates multiple signaling pathways in *Arabidopsis*. *Genes Dev.* **30**, 164–176.

Zentella, R., Sui, N., Barnhill, B., Hsieh, W.P., Hu, J., Shabanowitz, J., Boyce, M. *et al.* (2017) The *Arabidopsis* O-fucosyltransferase SPINDLY activates nuclear growth repressor DELLA. *Nat. Chem. Biol.* **13**, 479–485.

Zhu, Y. and Chen, X. (2017) Expanding the scope of metabolic glycan labeling in *Arabidopsis thaliana*. *Chembiochem* **18**, 1286–1296.

Zhu, Y., Wu, J. and Chen, X. (2016) Metabolic labeling and imaging of N-linked glycans in *Arabidopsis thaliana*. *Angew. Chem. Int. Ed. Engl.* **55**, 9301–9305.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Schematic of GalNAz and GlcNAz salvage pathways.
Figure S2 Gross morphologies of NP treated with GalNAz.
Figure S3 Structure of alkyne-AC-biotin.
Figure S4 Venn diagram of O-GlcNAcylation sites and O-GlcNAcylated proteins identified with different dissociation methods.
Figure S5 Motif of ambiguous O-GlcNAcylation sites identified by HCD-pd-EThcD.
Figure S6 Analysis of unambiguous O-GlcNAcylation sites in rice by different dissociation methods.
Figure S7 Functional analysis of O-GlcNAcylated proteins.
Figure S8 Analyses of proteins containing both O-GlcNAcylation and phosphorylation sites.
Figure S9 Evolutionary conservation and divergence of OGT between rice and humans.
Figure S10 Auto-O-GlcNAcylation of OsOGT2.
Figure S11 Validation of rice O-GlcNAcylated proteins in *Escherichia coli*.
Data S1 O-GlcNAylated proteins and sites identified by ETD.
Data S2 O-GlcNAylated proteins and sites identified by EThcD.
Data S3 O-GlcNAylated proteins and sites identified by HCD-pd-EThcD.
Data S4 Summary of O-GlcNAcylated proteins.
Data S5 Summary of phosphorylated proteins.
Data S6 Proteins modified by both O-GlcNAcylation and phosphorylation.
Table S1 Hyper-O-GlcNAcylated proteins.
Table S2 Primer list.