JBC RESEARCH ARTICLE

In vivo mapping of the mouse *Galnt3*-specific O-glycoproteome

Received for publication, January 22, 2024, and in revised form, July 23, 2024 Published, Papers in Press, August 2, 2024,
https://doi.org/10.1016/j.jbc.2024.107628

Kruti Dalal[1], Weiming Yang[1], E. Tian[2], Aliona Chernish[1], Peggy McCluggage[1], Alexander J. Lara[1], Kelly G. Ten Hagen[2], and Lawrence A. Tabak[1,*]

From the [1]Biological Chemistry Section and [2]Developmental Glycobiology Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA

Reviewed by members of the JBC Editorial Board. Edited by Robert Haltiwanger

The UDP-N-acetylgalactosamine polypeptide:N-acetylgalactosaminyltransferase (GalNAc-T) family of enzymes initiates O-linked glycosylation by catalyzing the addition of the first GalNAc sugar to serine or threonine on proteins destined to be membrane-bound or secreted. Defects in individual isoforms of the GalNAc-T family can lead to certain congenital disorders of glycosylation (CDG). The polypeptide N-acetylgalactosaminyltransferase 3 (GALNT)3-CDG, is caused by mutations in *GALNT3*, resulting in hyperphosphatemic familial tumoral calcinosis due to impaired glycosylation of the phosphate-regulating hormone fibroblast growth factor 23 (FGF23) within osteocytes of the bone. Patients with hyperphosphatemia present altered bone density, abnormal tooth structure, and calcified masses throughout the body. It is therefore important to identify all potential substrates of GalNAc-T3 throughout the body to understand the complex disease phenotypes. Here, we compared the *Galnt3*$^{-/-}$ mouse model, which partially phenocopies GALNT3-CDG, with WT mice and used a multicomponent approach using chemoenzymatic conditions, a product-dependent method constructed using EThcD triggered scans in a mass spectrometry workflow, quantitative O-glycoproteomics, and global proteomics to identify 663 *Galnt3*-specific O-glycosites from 269 glycoproteins across multiple tissues. Consistent with the mouse and human phenotypes, functional networks of glycoproteins that contain GalNAc-T3-specific O-glycosites involved in skeletal morphology, mineral level maintenance, and hemostasis were identified. This library of *in vivo* GalNAc-T3-specific substrate proteins and O-glycosites will serve as a valuable resource to understand the functional implications of O-glycosylation and to unravel the underlying causes of complex human GALNT3-CDG phenotypes.

A family of enzymes, the UDP-GalNAc polypeptide:N-acetylgalactosaminyltransferases (GalNAc-Ts), initiates the addition of the first sugar in mucin-type O-linked glycosylation by catalyzing the transfer of N-acetylgalactosamine (GalNAc) from the sugar donor UDP-GalNAc to the hydroxyl group of threonine/serine residues of the protein core (1). The GalNAc-T family of enzymes is composed of multiple members (20 in humans and 19 in mice), and these isoforms display both distinct but sometimes overlapping *in vitro* substrate specificities, suggesting that at least some isoforms may be responsible for the glycosylation of specific sites within proteins *in vivo* (1, 2).

Congenital disorders of glycosylation (CDG) result from genetic defects in protein and lipid glycosylation (3). Mutations in *GALNT3* lead to polypeptide N-acetylgalactosaminyltransferase 3 (GALNT)3-CDG, an autosomal recessive disorder of O-glycosylation, first reported by Topaz *et al.* in 2004 (4) and characterized by hyperphosphatemia and ectopic calcifications throughout the body. The GALNT3-CDG, is the most prevalent genetic cause of hyperphosphatemic familial tumoral calcinosis (HFTC) in humans. The absence of functional GalNAc-T3 activity results in the failure to glycosylate the phosphate-regulating hormone fibroblast growth factor 23 (FGF23) within osteocytes of the bone. The absence of a "shielding" O-glycan renders FGF23 susceptible to inactivating cleavage, with resultant dysregulation of phosphate homeostasis. A mouse model lacking the *Galnt3* gene partially phenocopies GALNT3-CDG, and these animals develop hyperphosphatemia without calcification (5), although ectopic calcification develops on a high phosphate diet (4, 6, 7). The Consortium for Functional Glycomics phenotype survey of *Galnt3*$^{-/-}$ mice also revealed decreased levels of monocytes and platelets that might influence chronic inflammatory reactions and may lead to defects in coagulation (8). Peluso *et al.* reported that the absence of *Galnt3* resulted in altered glycosylation of major murine salivary mucin and changes in the structure, composition, and stability of the oral microbiome, suggesting that the lack of *Galnt3* could result in oral disease (9). Therefore, the loss of *Galnt3* across multiple tissues where it is normally expressed likely underlies the complex presentations and phenotypic variability seen in both mice and human patients. To fully understand complex phenotypes such as these, it is necessary to identify all potential substrates of this glycosyltransferase.

While *GALNT3* has a more restricted expression pattern than the ubiquitously distributed *GALNT1* and *GALNT2*, it is still found in many tissues throughout the body (10). Therefore, GalNAc-T3 may influence phosphate homeostasis not only by affecting the activity of FGF23 signaling, but may directly regulate other substrate proteins involved in phosphate and calcium homeostasis, such as those found in colon,

* For correspondence: Lawrence A. Tabak, lawrence.tabak@nih.gov.

kidneys, and submandibular and sublingual salivary glands (SGs). To gain a more comprehensive understanding of how the loss of *Galnt3* results in disease, we used quantitative O-glycoproteomics and global proteomics (11) to identify GalNAc-T3 specific solitary O-glycosites *in vivo* across multiple tissues by comparing WT and *Galnt3*$^{-/-}$ mice. We identified a total of 1197 O-glycosites from 414 glycoproteins in *Galnt3*$^{-/-}$ and WT mice, with 663 O-glycosites from 269 glycoproteins that appear to be *Galnt3*-specific sites. Many of these sites were found in glycoproteins expressed in the submandibular gland (SMG) and the sublingual gland (SLG). Several GalNAc-T3 sites were found on extracellular matrix (ECM) proteins, such as aggrecan, versican, brevican, and collagen alpha-1 chain in SGs and other tissues, which play important roles in the development of the skeletal system. We also identified substrates involved in the complex process of hemostasis, including Von Willebrand factor, platelet glycoprotein Ib alpha chain, coagulation factor V, and kininogen-1.

This study represents an important step toward unraveling the landscape of O-GalNAc glycosylation specific to GalNAc-T3 and establishes a strong groundwork for decoding the underlying causes of complex human GALNT3-CDG phenotypes.

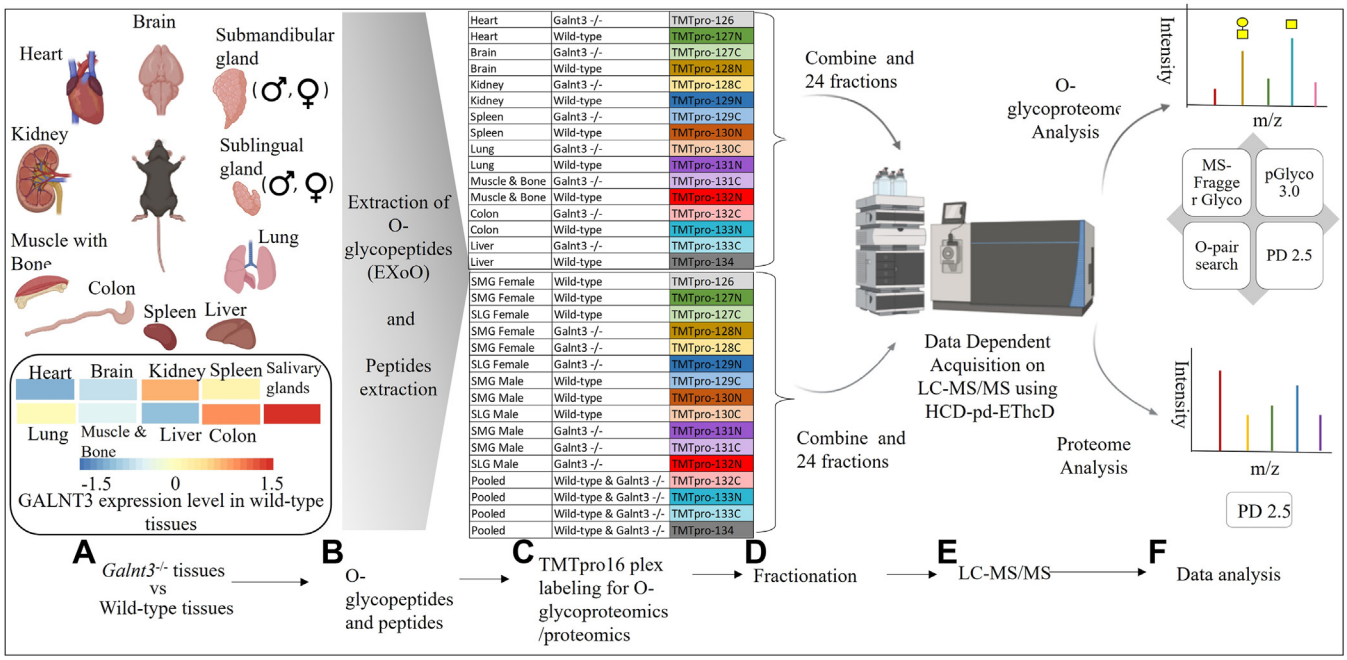
Results

Study design for identification of Galnt3*-specific sites*

We evaluated the consequences of the systemic loss of GalNAc-T3 across multiple tissues by comparison of *WT* and *Galnt3*$^{-/-}$ mice. The workflow is shown in Figure 1. We used a method, termed site-specific extraction of O-linked glycopeptides (EXoO), to identify solitary O-glycosites by cleaving the N-termini of O-glycosylated serines and threonines in an O-glycan-dependent way with the O-glycoprotease, OgpA (12, 13). This presents the O-glycosites as the first amino acid in the identified peptide sequences and enhances confidence in the assignment of O-glycosites (13). All products were detected using a product ion-triggered approach (HCD-pd-EThcD). The quantitative glycoproteomics and proteomics analysis of tissues from *Galnt3*$^{-/-}$ and *WT* mice was performed using TMTpro16 plex labeling. Glycopeptide search engines—MSFragger-Glyco (https://msfragger.nesvilab.org/) (14), pGlyco3.0 (https://github.com/pFindStudio/pGlyco3) (15), and O-pair Search with MetaMorpheus (https://github.com/smith-chem-wisc/MetaMorpheus) (16) were used to localize O-glycosites (Fig. S1) and their corresponding O-glycoproteins with glycan-level accuracy and deduce those sites which are *Galnt3*-specific. Using quantitative O-glycoproteomics and proteomics approaches, we generated ten tissue maps of GalNAc-T3 specific O-glycosites, the majority of which have not been previously reported. Our datasets identified 663 O-glycosites corresponding to 269 glycoproteins with reduced O-glycopeptide intensity and no change in protein abundance in null mice.

The comparison of the protein expression profiling between WT *and* Galnt3$^{-/-}$ *mice*

In the global proteomics analysis, a total of 4884 proteins were identified in eight tissues (Dataset S1) and 4062 proteins



Heart	Galnt3 -/-	TMTpro-126
Heart	Wild-type	TMTpro-127N
Brain	Galnt3 -/-	TMTpro-127C
Brain	Wild-type	TMTpro-128N
Kidney	Galnt3 -/-	TMTpro-128C
Kidney	Wild-type	TMTpro-129N
Spleen	Galnt3 -/-	TMTpro-129C
Spleen	Wild-type	TMTpro-130N
Lung	Galnt3 -/-	TMTpro-130C
Lung	Wild-type	TMTpro-131N
Muscle & Bone	Galnt3 -/-	TMTpro-131C
Muscle & Bone	Wild-type	TMTpro-132N
Colon	Galnt3 -/-	TMTpro-132C
Colon	Wild-type	TMTpro-133N
Liver	Galnt3 -/-	TMTpro-133C
Liver	Wild-type	TMTpro-134
SMG Female	Wild-type	TMTpro-126
SMG Female	Wild-type	TMTpro-127N
SLG Female	Wild-type	TMTpro-127C
SMG Female	Galnt3 -/-	TMTpro-128N
SMG Female	Galnt3 -/-	TMTpro-128C
SLG Female	Galnt3 -/-	TMTpro-129N
SMG Male	Wild-type	TMTpro-129C
SMG Male	Wild-type	TMTpro-130N
SLG Male	Wild-type	TMTpro-130C
SMG Male	Galnt3 -/-	TMTpro-131N
SMG Male	Galnt3 -/-	TMTpro-131C
SLG Male	Galnt3 -/-	TMTpro-132N
Pooled	Wild-type & Galnt3 -/-	TMTpro-132C
Pooled	Wild-type & Galnt3 -/-	TMTpro-133N
Pooled	Wild-type & Galnt3 -/-	TMTpro-133C
Pooled	Wild-type & Galnt3 -/-	TMTpro-134

Figure 1. Scheme of workflow. *A*, the tissues were dissected from C57BL/6NHsd (WT) and *Galnt3*-null (*Galnt3*$^{-/-}$) mice. A heatmap of the relative abundance of GalNAc-T3 in tissues reported from https://www.proteomicsdb.org/ is shown. Due to the high abundance of GalNAc-T3 in both SMGs and SLGs and well-established sexual differences in SMGs in rodents, SMG and SLG from male and female mice were dissected and analyzed separately. *B*, O-glycopeptides and peptides were extracted from tissues using the extraction of O-glycopeptides (ExoO) method. *C*, TMTpro16-plex, labeled purified O-glycopeptides and peptides from each tissues derived from WT and *Galnt3*$^{-/-}$ mice. *D*, labeled samples were combined into 24 fractions using high-pH reversed-phase liquid chromatography. *E*, data were acquired on LC-MS/MS with HCD-pd-EThcD. *F*, O-glycoproteomics data were analyzed using MSFragger, pGlyco3.0, and O-pair search with Metamorpheus and proteomics data, and reporter ion intensity of TMT tags was analyzed using PD 2.5. GALNT, polypeptide N-acetylgalactosaminyltransferase 3; HCD, higher-energy collisional dissociation; SLG, sublingual gland; SMG, submandibular gland; TMT, tandem mass tag.

in SMGs and SLGs (Dataset S2). SMG (containing both mucous and serous acinar cells) and SLG (containing mainly mucous acinar cells) each produce mucin-rich secretions. Mucin levels in SLGs are ten-fold higher than in SMGs and these glands produce a more viscous secretion. Muc19 is synthesized and secreted exclusively by the SLGs (17) and Muc10, is produced by the SMGs (18). In addition, Muc2, Muc5Ac, and Muc6 are specifically found in SLGs, while Muc1, Muc7, Muc13, and Muc18 are found in both murine SMGs and SLGs. Among the identified proteins, 3262 were identified in all tissues, while 800 proteins were specific to the SMGs and SLGs and 1622 proteins were found in the other tissues examined. To understand the profile of O-glycosites in various tissues, the average relative abundance of murine Galnt proteins and transcripts were obtained from the proteomicsDB. Overall, the transcripts and protein levels do not correlate well (Figs. 2*A* and S2). The unsupervised hierarchical cluster analysis of Galnt proteins revealed that the brain exhibits the most distinct expression profile (Fig. 2*A*). The heart and kidney are closely grouped, while the spleen and colon exhibit similar patterns. Additionally, liver, lung, SMGs, SLGs, and muscle with bone are clustered together due to similarities in GalNAc-Ts expression as shown in Figure 2*A*. Similar to the data from ProteomicsDB presented in Figure 2*B*, our quantitative proteomics data showed a similar pattern of GalNAc-Ts expression in WT, with GalNAc-T2 being strongly expressed in all tissue types and GalNAc-T11 having higher expression in the kidney, colon, and brain. Our quantitative proteomics tissue data did not reveal the presence of additional GalNAc-Ts. The ProteomicsDB data did not provide distinct intensity values for individual SMGs and SLGs; instead, they were grouped together as "salivary glands". In Figure 2*A*, the heatmap illustrated that GalNAc-T7 and T3 exhibited higher expressions in the combined SMGs and SLGs, followed by T18 = T10. Additionally, T2, T12, T4, T1, T5, and T6 showed decreasing levels of expression. Based on our quantitative proteomics analysis of the SG data, it was observed that both male and female SLGs exhibited elevated expression of GalNAc-Ts compared to SMGs (Fig. 2*C*). In SLG of both sexes, GalNAc-T7 demonstrated the highest expression, followed by T12, T5, T4, T3, T2, and T10 in descending order of abundance, while in the SMG, T7 was also prominently expressed, with T3 following closely in terms of abundance.

The absence of *Galnt3* led to significant variations in protein expression levels across different tissues, according to the findings outlined in Figure 2, *D* and *E*. This variation ranged from 663 proteins with altered levels in female SLG to just eight proteins in the brain. The absence of *Galnt3* had a significant impact on O-glycosylation, with varying effects in different tissues. Among the colon, kidney, spleen, and lung which exhibited high to low levels of GalNAc-T3 expression in that order, 46, 37, 36, and 27 substrates were identified, respectively (Dataset S3). The SMGs and SLGs, in both males and females, showed the highest number of glycosites specific to the GalNAc-T3 (Fig. 3*A* and Dataset S4). Changes in the levels of specific proteins suggest that GalNAc-T3 has diverse roles across different tissues. For example, in the heart derived from *Galnt3*$^{-/-}$ mice, we identified decreased levels of glycoprotein V, platelet glycoprotein Ib alpha chain, and Trem-like transcript 1 protein, which are associated with platelet function. In the SLG derived from male *Galnt3*$^{-/-}$ mice, we observed increased abundance of kallikrein-1, kallikrein 1-related peptidases, and sodium/potassium-transporting ATPase subunits, which could affect the endocrine and other factor-regulated calcium reabsorption pathway. In the absence of *Galnt3*, the liver displayed a reduction in the abundance of 114 proteins. Among them, tartrate-resistant acid phosphatase 5b (Acp5), an enzyme known to be involved in the regulation of bone resorption (19) may influence bone volume and the activity of osteoclasts observed in *Galnt3*$^{-/-}$ mice. In the kidney from *Galnt3*$^{-/-}$ mice, we observed a decrease in the protein abundance of regucalcin, which is involved in calcium homeostasis (20) potentially influencing calcium regulation in the kidney. Altered proteins in the lung of *Galnt3*$^{-/-}$ mice included those involved in complement and coagulation cascades, hemostasis, fibrinolysis, and regulation of platelet activation. An increase was observed in alpha-2-antiplasmin protein in *Galnt3*$^{-/-}$ mice which plays an important role in the regulation of blood coagulation. These findings indicate that the absence of *Galnt3* has diverse impacts on various physiological aspects in different tissues, contributing to the maintenance of tissue-specific functions.

Quantitative O-glycoproteomics and proteomics identify GalNAc-T3-specific substrate

O-glycosites from the different mouse tissues were mapped using MSFragger-Glyco, pGlyco3.0, and O-pair. Figure 3, *B–E*) displays an Euler diagram of the glycopeptide sequences, glycopeptides, glycoproteins, and glycan compositions independently detected by the three software programs. Within the glycopeptides identified by three search engines, EThcD confirmed O-glycosites being at the first amino acid position of the glycopeptide sequences, and tandem mass tag (TMT) reporter ion abundance values were used for downstream analysis to identify substrates for GalNAc-T3 by comparing *Galnt3*$^{-/-}$ and WT tissues.

We identified 1197 O-glycosites from 414 glycoproteins in samples from *Galnt3–/–* and WT mice. Of these, 711 O-glycosites from 291 glycoproteins showed at least a 1.5-fold reduction in tissues derived from *Galnt3*$^{-/-}$ mice, suggesting they were potential GalNAc-T3 substrates. *Galnt3* null effects on O-glycosite abundance across tissues are displayed as volcano plots (Fig. 3*A*) and summarized in the bar charts (Fig. 4*A*). The decreased abundance of O-glycosites could be either an effect of loss of GalNAc-T3 activity or a reduced abundance of glycoproteins in the null mice. While we were unable to determine the abundance of 204 glycoproteins in the global proteome data, a total of 22 glycoproteins were decreased in abundance across different tissues that were associated with 42 O-glycosites displaying reduced abundance as shown in Figure 4*B*. A total of 669 O-glycosites from 269 glycoproteins were decreased in tissues derived from *Galnt3*$^{-/-}$mice, with no change in protein levels in the global

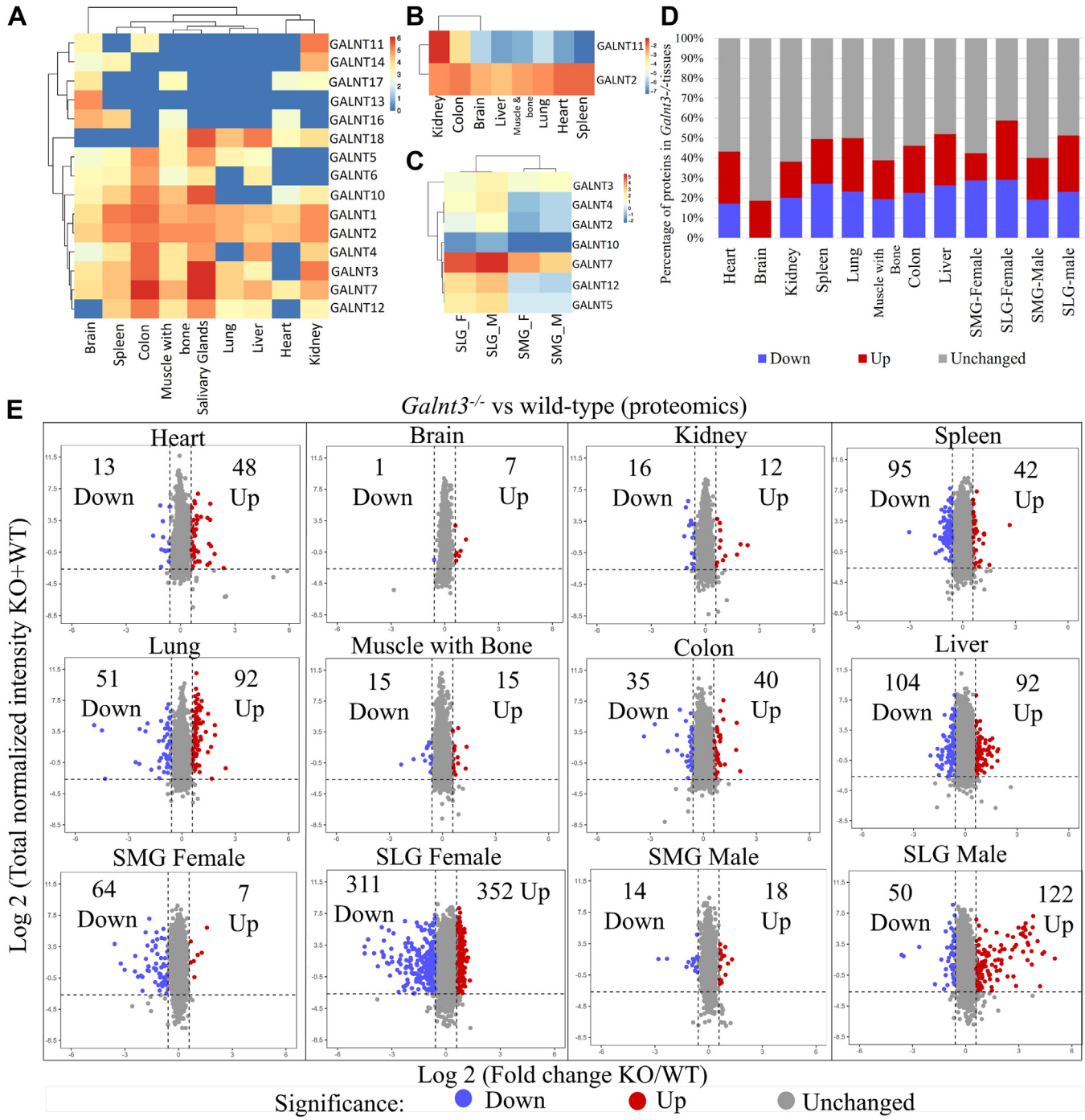


Figure 2. Expression of proteins in *Galnt3*$^{-/-}$ *versus* WT tissues. *A*, heatmap, unsupervised hierarchical clustering, of relative abundance of GalNAc-Ts in different tissues reported from https://www.proteomicsdb.org/. *B*, heatmap showing protein abundance of GalNAc-T2 and GalNAc-T11 across different WT tissues from our quantitative proteomics data. *C*, protein abundance of GalNAcT-2,3,4,5,7,10,12 from our proteomics data for male and female salivary glands. *D*, bar graphs show the percentage of proteins with altered abundance levels in *Galnt3*$^{-/-}$ tissues relative to WT. *E*, protein abundance ratios of *Galnt3*$^{-/-}$/WT across tissues, are represented as volcano plots. Each *dot* represents a protein. *Gray*, *blue*, and *red* dots represent unchanged, decreased (foldchange <1.5) and increased (foldchange >1.5) protein abundance, respectively, in the absence of *Galnt3*. GALNT, polypeptide N-acetylgalactosaminyltransferase 3.

proteomics data, indicating that they could be GalNAc-T3 substrates. Figure 4*C* shows the identification of O-glycosites specific to GalNAc-T3 in more than three tissues. We identified 203 O-glycosites from 51 glycoproteins in numerous cell adhesion molecules across different tissues, with no alteration at the protein level. The core proteins of versican, neurocan, aggrecan, and brevican, known to carry mucin-type O-glycosylation in humans (13, 21–23), have been found to have several GalNAc-T3 sites. GalNAc-T3-specific sites were also found on osteocalcin (Bglap, S57), osteocalcin-related protein (Bglap3, S57 and S64), C-type lectin domain family 11 member A (Clec11a, T108) and the putative sodium-coupled neutral amino acid transporter 10 (Slc38a10, S1047, S429, S636, S819, T769, and T821), all of which are involved in

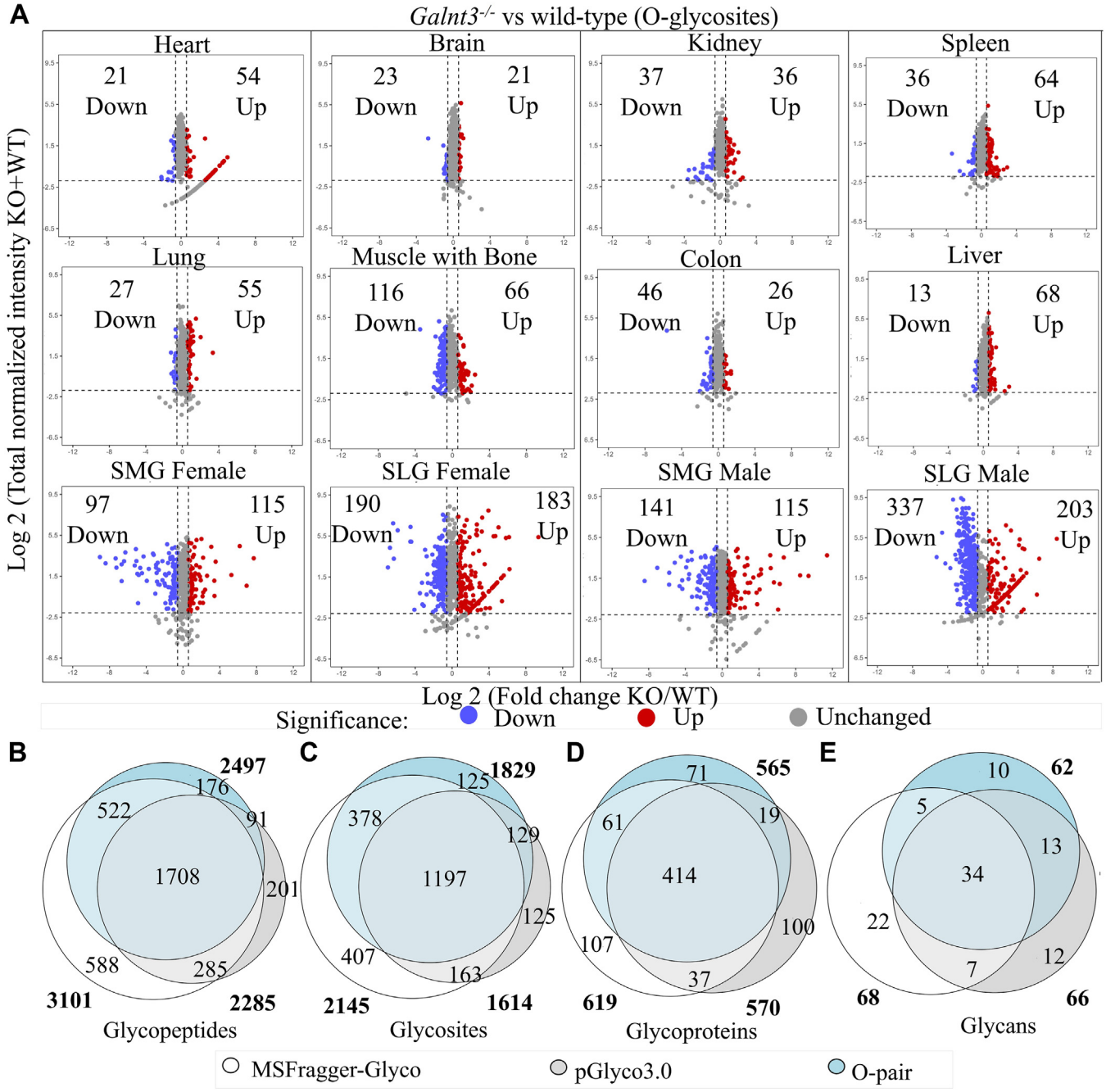


Figure 3. Mapping of O-glycosites from different tissues. *A*, *Galnt3*$^{-/-}$ effects on O-glycosite abundance across tissues, represented as volcano plots. Each dot represents a glycosite. *Gray*, *blue*, and *red dots* represent unchanged, decreased (foldchange <1.5), and increased (foldchange >1.5) O-glycosite abundance, respectively. *B–E*, euler diagrams of glycopeptide sequences, glycopeptides, glycoproteins, and glycan compositions separately identified by MSFragger-Glyco, pGlyco3.0, and O-pair. The area proportional Euler diagram represents sets and their relationships, illustrating the common and distinct O-glycopeptides/peptides/glycans identified using MSFragger, O-pair, and pGlyco3.0. GALNT, polypeptide N-acetylgalactosaminyltransferase 3.

bone structure and ossification (24–26). In addition, *Galnt3* specific O-glycosites were mapped to coagulation factor V (F5, S1130, S1488, T1111, T1192, T1342, T1350, T1369, and T1378), Von Willebrand factor (vwf, S1254, S1255, T1468, and T743) and platelet glycoprotein Ib alpha chain (Gp1ba, T358). A detailed list of GalNAc-T3 specific substrates across the tissues is summarized in the Dataset S5. It is worth noting that we also observed an increase in certain O-glycosites in the *Galnt3*$^{-/-}$ tissues as shown in Figure 3*A*. This increase could potentially be attributed to other Galnts, which might be acting to compensate for the loss of GalNAc-T3. However, the precise mechanism behind this phenomenon remains unclear, and it would be intriguing to investigate the impact the ablation of one isoform on the function of the remaining isoforms.

We identified O-glycosites with reduced levels of glycosylation, and we were uncertain whether this reduction was a consequence of the absence of *Galnt3* or if it was due to overall decreased levels of the proteins themselves. Specifically,

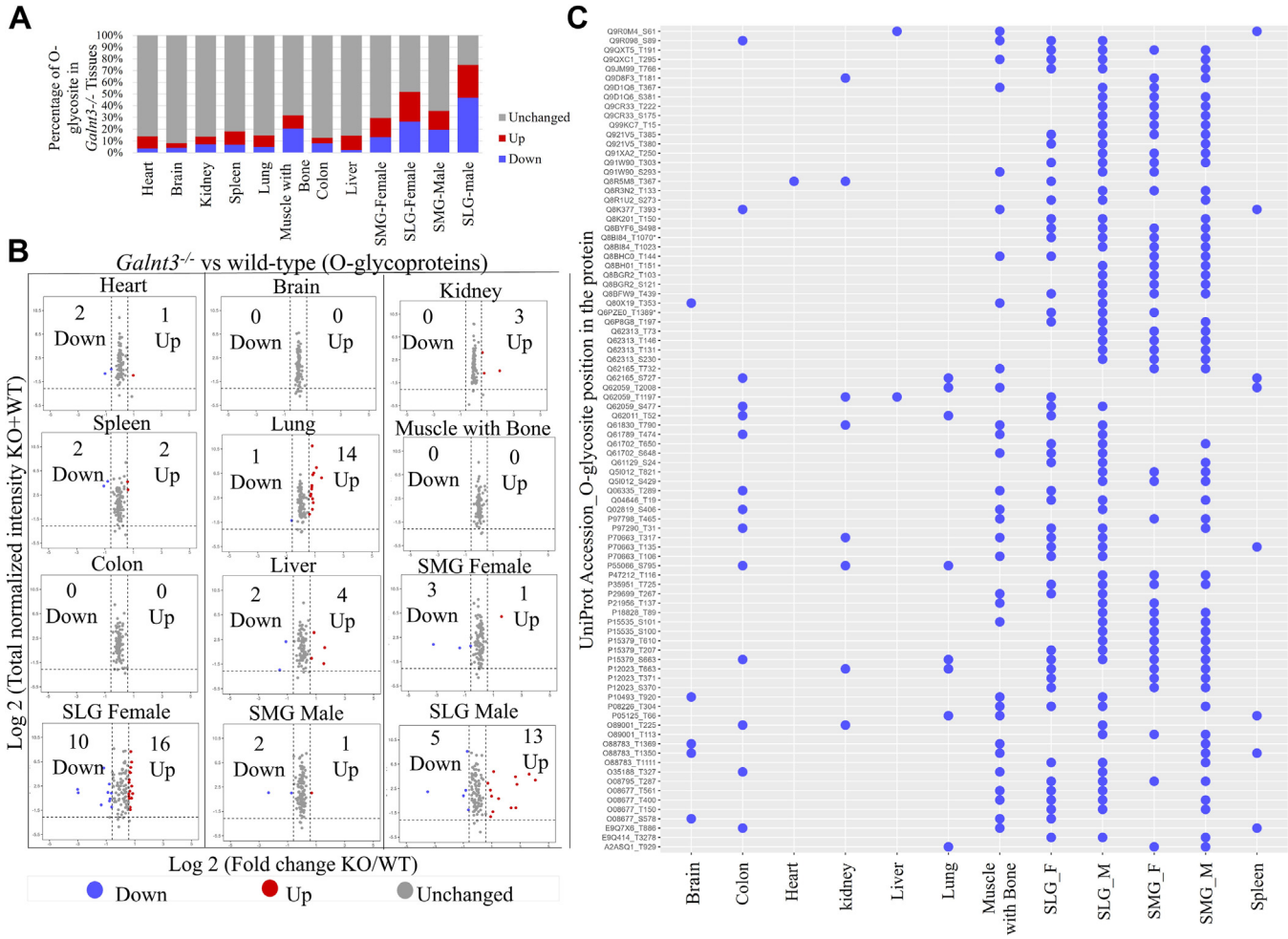


Figure 4. Catalog of GalNAcT-3 specific O-glycosites across tissues. *A*, the bar graph shows the percentage of unchanged and changed O-glycosites in tissues dissected from *Galnt3*$^{-/-}$mice. *B*, glycoprotein abundance across tissues, represented as volcano plots. Each *dot* represents a glycoprotein that corresponds to GalNAc-T3-specific O-glycosites. *Gray*, *blue*, and *red* dots represent unchanged, decreased (foldchange <1.5), and increased (foldchange >1.5) glycoprotein abundance, respectively. *C*, dot plot of GalNAc-T3 isoform specific O-glycosites identified in more than three tissue types. Each *blue dot* represents a GalNAc-T3 substrate. The *x*-axis represents the tissue types and the *y*-axis represents the GalNAc-T3 substrate (UniProtAccession_O-glycosite position in the protein). Q8BI84_T1070* glycosite was identified at T1086; T1102 and Q6PZE0_T1389* had multiple sites at T1552; T1715; T1878; T2041; T2204; T2367; T2530; T2693; T2856; T3019; T3182; T3345; T3508; T3671; T3834; T3997; T4160; T4323; T4486; T4649; T4812; T4975; T5138; T5301; T5464; T5627; T5790; T5953; T6116; T6279; T6442; T6605; T6768; T6931; and T7094. GALNT, polypeptide N-acetylgalactosaminyltransferase 3.

we found that 24 O-glycosites, localized to 16 glycoproteins, showed decreased levels in tissues lacking *Galnt3* compared to the *WT* samples. Notably, some specific glycoproteins, such as platelet glycoprotein Ib alpha chain (S366 and T358), Trem-like transcript 1 protein (T147), thioredoxin domain-containing protein 5 (S293), microfibrillar-associated protein 5 (T39), and selenoprotein P (T235), were reduced in tissues derived from *Galnt3*$^{-/-}$ mice. Moreover, in the SGs, we observed a decrease in 18 GalNAc-T3-specific O-glycosites, corresponding to 11 distinct proteins that showed reduced levels. Examples of these proteins include lactadherin (T129 and T137), collagen alpha-1(XIV) chain (T353 and T1011), and sulfhydryl oxidase 1 (T282 and T652), among others.

The majority of the GalNAc-T3 specific O-glycosites were found in SMGs and SLGs

As seen in Figure 5*A*, both SGs and sex-specific O-glycosites were identified, where each bar represent the O-glycosite identified. In Figure 5*B*, a total of 94 and 139 GalNAc-T3 specific O-glycosites were identified, out of which, 30 and 75 O-glycosites were unique to female and male SMG, respectively. Growth factors such as, beta-nerve growth factor (T56 and T96) and proepidermal growth factor (S825 and T826) were identified as GalNAc-T3 substrates in the male SMG. Our quantitative proteomics data revealed that these growth factors are produced in larger quantities in male SMGs than in females. This difference in production could be attributed to their synthesis in the tubular duct cells and storage in the granules of tubular duct cells, which are larger than acinar cells in males (27). The intensity of GalNAc-T3, one of the most abundant isoforms in SMGs and SLGs, was reduced in both SMG and SLG in both sexes by at least 2.3 Log2 fold change in glands derived from *Galnt3*$^{-/-}$ mice using quantitative proteomics. We identified a total of 488 GalNAc-T3 specific O-glycosites from 215 O-glycoproteins from SGs. In SMG derived from *Galnt3*$^{-/-}$ mice, the abundance of 94 sites in

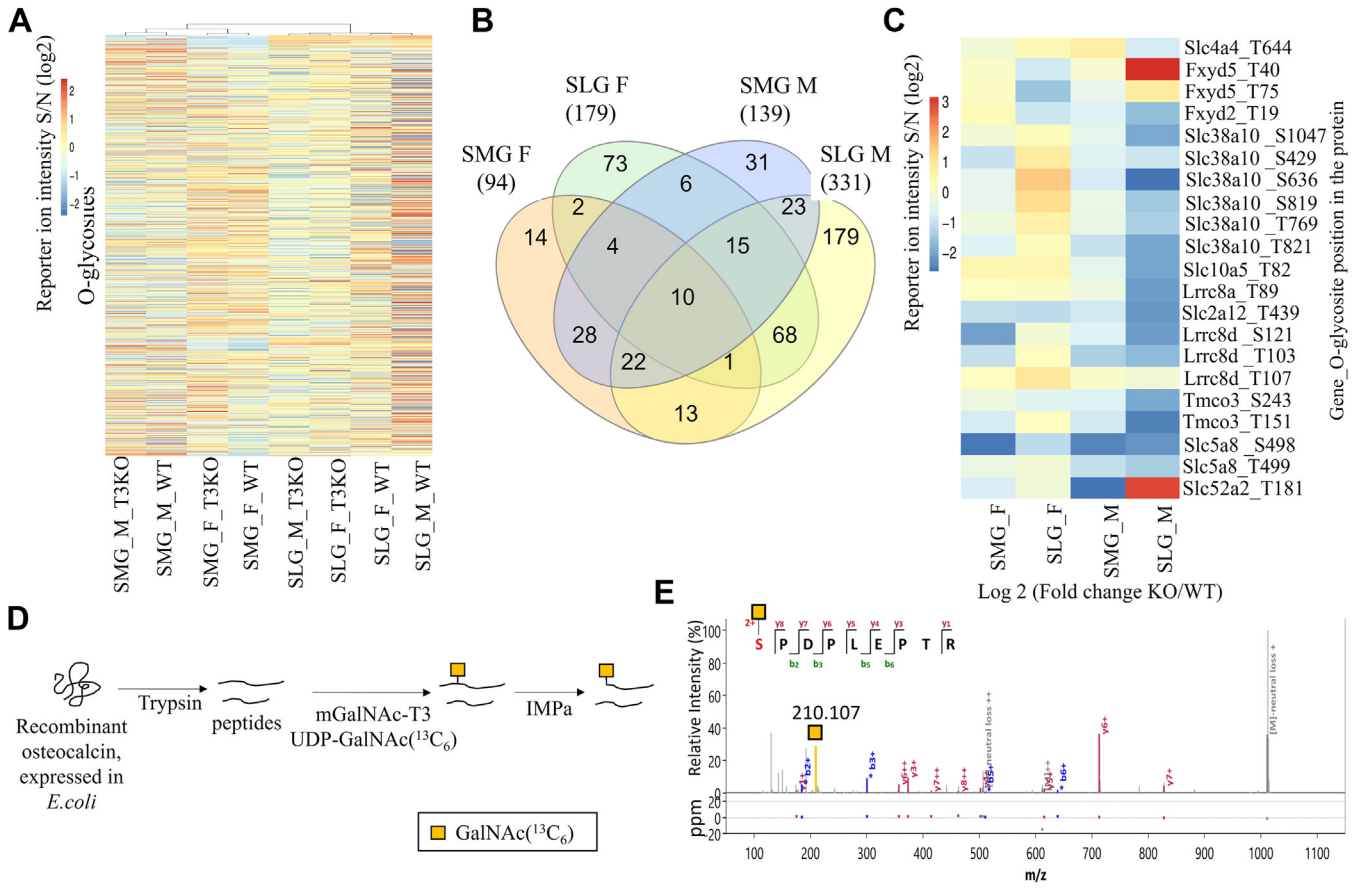


Figure 5. Mapping of O-glycosites in SMGs and SLG. *A*, heatmap of O-glycosites identified in SMGs and SLGs from both sexes from *Galnt3*$^{-/-}$ *versus* WT. *B*, Venn diagram of GalNac-T3 substrates identified in the salivary glands based on sex. *C*, heatmap of GalNAc-T3 specific sites identified in the transporters category. *D*, *In vitro* glycosylation workflow to validate osteocalcin (Bglap) S57 as a GalNAc-T3-specific substrate. *E*, MS2 spectrum of S57 site on recombinant mouse osteocalcin (Bglap) glycosylated by murine GalNAc-T3 and GalNAc ($^{13}C_6$) followed by IMPa cleavage. GALNT, polypeptide N-acetylgalactosaminyltransferase 3; IMPa, immunomodulating metalloprotease; SMG, submandibular gland; SLG, sublingual gland.

females and 139 sites in males was decreased relative to *WT* (Fig. 5*B*). In SLGs derived from *Galnt3*$^{-/-}$ mice, 179 sites in females and 331 sites in males were decreased relative to WT (Fig. 5*B*). A total of 40 glycosites from mucin-19 were identified as substrates for GalNAc-T3 in the SLG, where the SAPTASSTAAGSGLR site was identified as a substrate at the multiple locations in this gel secreting mucin. Mucin-like protein2 (T32, T95, and T105), endomucin, (S43), and mucin-13 (T231 and S33) were identified as GalNAc-T3 specific substrates in the SGs. In addition, several GalNAc-T3 specific sites were identified in transporters such as solute carrier family 2, facilitated glucose transporter member 12 (slc2a12 and T439), sodium-coupled monocarboxylate transporter 1 (Slc5a8; T499, and S498), and the regulatory unit of transporters such as sodium/potassium-transporting ATPase subunit gamma (fxyd2; T19) as shown in Figure 5*C*. Osteocalcin (Bglap) S57 was validated *in vitro* as a GalNAc-T3 substrate (Fig. 5*D*). After the digestion with IMPa, a GalNAc($^{13}C_6$) labeled glycopeptide with S57 at the first amino acid position was detected, suggesting that GalNAc-T3 glycosylates Bglap S57 (Fig. 5*E*).

Additionally, trans-Golgi network integral membrane protein 1/2 (Tgoln1/2) was purified from male SLGs derived from WT and *Galnt3*$^{-/-}$ and subset of O-glycosites are validated (Table S1 and Dataset S6). Among the eight peptides tested *in vitro*, VATNPIITPVAQENQ and GIEDKITTTGATGGL peptides corresponding to apolipoprotein E and podoplanin showed higher glycosylation by hGalNAc-T3 compared to hGalNAc-T1 and hGalNAc-T2, supporting our glycoproteomic data for GalNAc-T3 selective substrates (Table S2 and Dataset S7).

GalNAc-T3 substrates map to various cellular components and distinct biological pathways

The 269 glycoproteins identified as GalNAc-T3 specific substrates were searched and analyzed for the Gene Ontology terms (Datasets S8–S10) to understand their potential function. The top 10 Gene Ontology terms ranked by enrichment false discovery rate (FDR) analysis are shown in Figure 6*A*. Substrate glycoproteins were enriched in biological processes including cell adhesion, ECM organization, receptor mediated endocytosis, and hemostasis (Dataset S8). As anticipated, the majority of the glycoproteins were enriched in the membrane, Golgi, extracellular region, and extracellular space (Dataset S9). Molecular functions included cargo receptor activity,

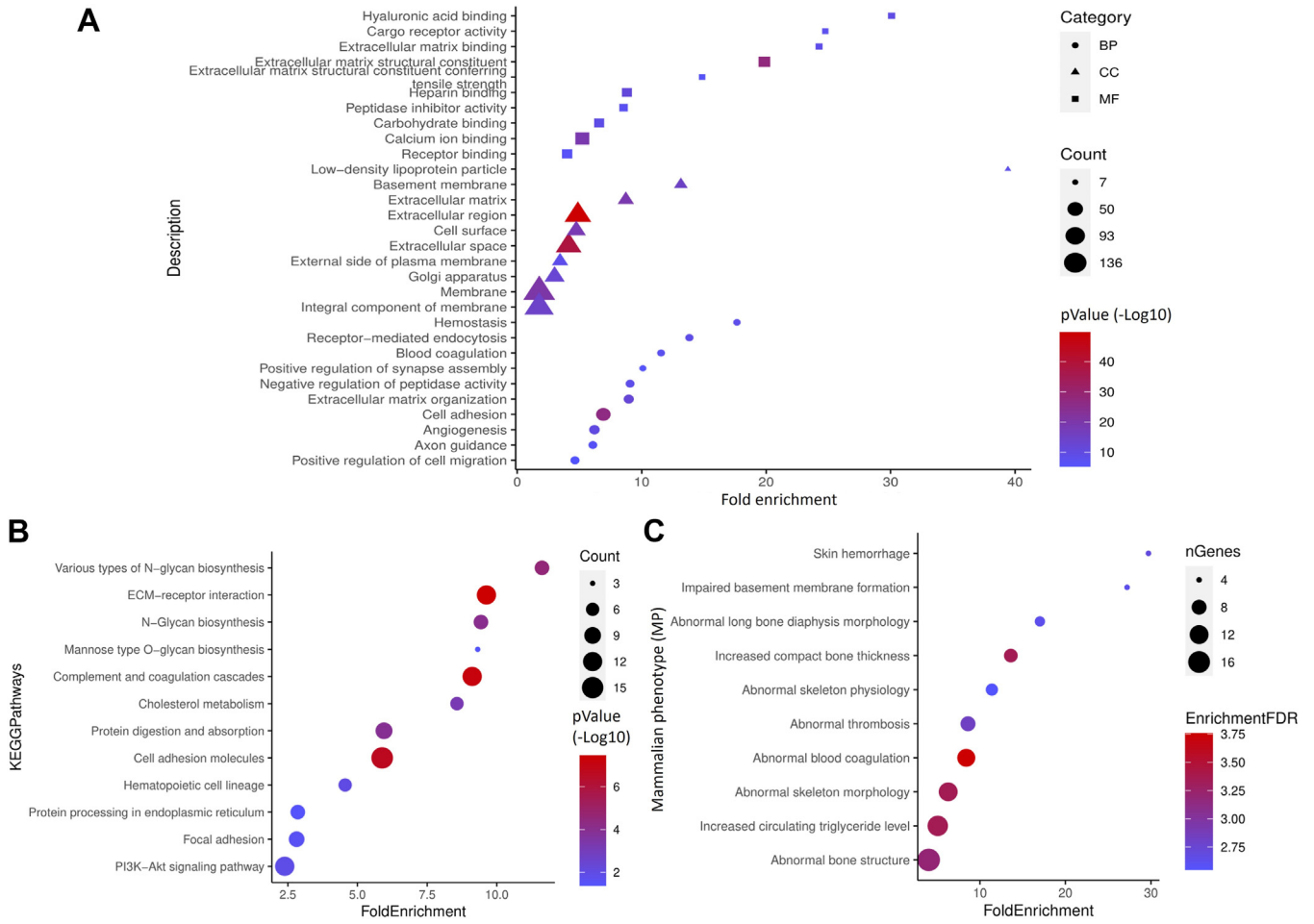


Figure 6. Analysis of glycoproteins corresponding to GalNAc-T3 substrates in biological pathways. (*A*) GO term analysis, (*B*) KEGG pathway analysis, (*C*) mammalian phenotype (MP) analysis of glycoproteins identified as GalNAc-T3 substrates. GE, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

ECM structural constituents and binding activities (Dataset S10). Kyoto Encyclopedia of Genes and Genomes pathway analysis (Fig. 6*B*) showed the involvement of glycoproteins in cell adhesion, ECM-receptor interaction and complement and coagulation cascades. Details about the glycoproteins that correspond to identified GalNAc-T3 substrates involved in Kyoto Encyclopedia of Genes and Genomes pathways are shown in Dataset S11. Next, the mammalian phenotype ontology tool (Mouse Genome Informatics) was used to describe the phenotypic data that corresponds to the identified GalNAc-T3 substrates (Fig. 6*C*). Interestingly, the mammalian phenotype tool identified pathways involved in abnormal bone structure, abnormal skeleton morphology, abnormal blood coagulation, increased circulating triglyceride levels, as well as impaired basement membrane formation (Fig. 6*C* and Dataset S12).

GALNT3-CDG phenotypic insights using STRING network analysis

The 269 glycoproteins identified as GalNAc-T3 specific substrates from tissues and SGs were analyzed using the STRING tool (version 12.0). With a high confidence (0.7) interaction score, the network's edge thickness reflects the strength and confidence of the data support (disconnected proteins in the network are not represented). The analysis revealed that GalNAc-T3 specific substrates across all tissues examined have a role in skeletal morphology, bone development and remodeling, mineral level maintenance, calcium ion binding, blood coagulation, and wound healing (Fig. 7). The network broadly depicts the common features of GALNT3-CDG including skeletal abnormalities, bone deformities, and disturbances in mineral metabolism.

The preferred substrate sequence for GalNAc-T3 using consensus motif analysis

Consensus motifs were determined for the 663 O-glycosites that were experimentally identified GalNAc-T3-specific substrates (Fig. 8). Sequences were aligned for seven residues N- and C-terminal from the OgpA cleavage sites (corresponding to O-glycosites) from identified O-glycopeptides. A total of 29 O-glycosites were not sequence aligned for ±7 residues as they were located either near the N or C termini of proteins. As shown in the Figure 8*A*, threonine and serine accounted for 475 (71.6%) and 314 (28.4%) of the O-glycosylation sites,

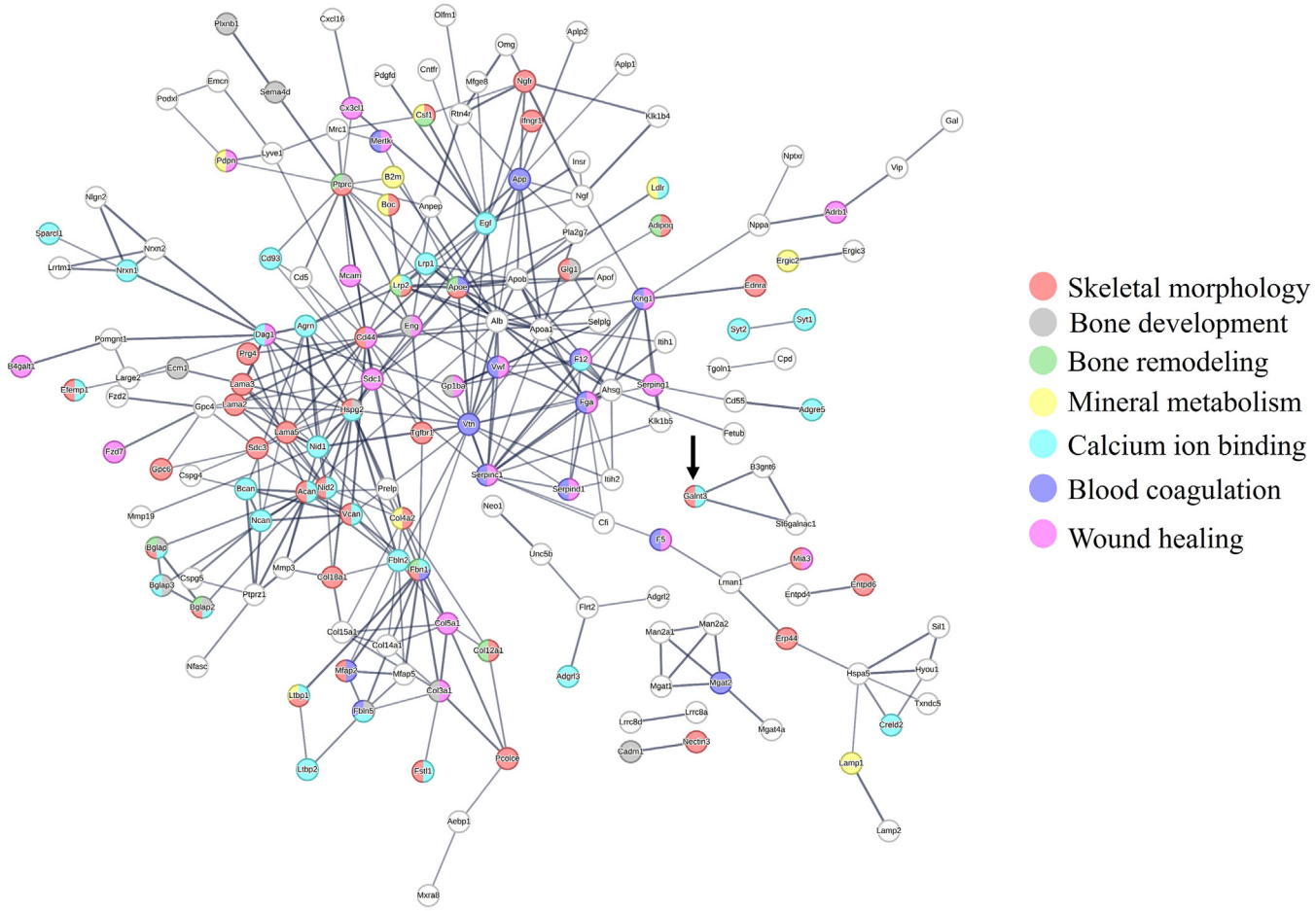


Figure 7. STRING network of glycoproteins identified as GalNAc-T3 substrates. Shown are all proteins identified as GalNAcT-3 substrates across all tissues examined. *Lines* indicate the strength and confidence of the data support. The *arrow* indicates the GalNAc-T3 protein which was included in the analysis. Extracellular matrix (ECM) remodeling and proteins (highlighted in *red nodes*) including laminins (lama2, lama3, and lama3), collagens (col18a1 and col4a2), proteoglycan core proteins (Vcan, Acan, Bcan, and Ncan) appear to a hub for protein-protein interactions in skeletal morphology pathway. Several proteins such as Vwf, F12, Serpinc1, Kng1, APP, Klk1, and Klk1b5 (highlighted in *blue* and *purple* nodes) are involved in complement and coagulation cascade. Interacting O-glycoproteins such as Eng, Ecm1, Cadm1, Fbln5, Sema4d, Plxnb1, Bglap, Col3a1, Slc38a10, Glg1, and Ptprc (highlighted in *green* and *gray* nodes) are involved in bone development and remodeling.

respectively. Preferred flanking amino acids at each position are shown for all sites (Fig. 8*A*), sites of threonine glycosylation (Fig. 8*B*) and sites of serine glycosylation (Fig. 8*C*). A comparison of the glycopeptide sequences surrounding the mapped O-glycosites revealed a preference for proline at position +3 (relative to the site of O-glycosylation) (29.71%) followed by position −1 (23.38%) and a high frequency of serine, threonine, alanine, and glutamate was favored near O-glycosites as depicted in the heatmap plot (Fig. 8, *D* and *E*). Next, a web-based tool ISOGlyP (Isoform-Specific O-Glycosylation Prediction) was used to predict the likelihood that a site would be glycosylated by GalNAc-T3. A total of 269 O-glycoproteins corresponding to 663 substrate O-glycosites were submitted to the ISOGlyP prediction algorithm. The ISOGlyP used positional residue specific enhancement values (EVs) spanning ± 5 residues from the serine or threonine glycosylation site and scored EV $>$ 1 (favored) or EV $<$ 1 (disfavored) in a peptide sequence glycosylated by a given GalNAc-T. A total of 35,097 O-glycosylation sites were predicted from 269 glycoproteins, out of which 3779 sites were predicted to be O-glycosylated by GalNAc-T3 (EV $>$ 1) (Fig. 8 *F* and *G*). After comparing experimentally identified sites and predicted sites, only 190 O-glycosites at threonine (Fig. 8, *F* and *H*) and one serine site DRAQVSPSPET from fibulin-2 (Fig. 8*F*) were predicted to be glycosylated by GalNAc-T3 isoform by IsoGlyP. Two hundred ninety threonine O-glycosites were not predicted by the algorithm (Fig. 8*I*). When analyzing flanking sequences of sites not identified by IsoGlyP, there was a notable presence of the acidic amino acids aspartate (D) and glutamate (E).

Discussion

Here, we provide the first comprehensive identification of *in vivo Galnt3* substrates across multiple mouse tissues. To better understand the phenotypic presentation in GALNT3-CDG, we mapped 663 O-glycosites from 269 glycoproteins across various tissues, including heart, brain, kidney, spleen, lung, muscle with bone, colon, liver, SMGs, and SLGs through comparison of *WT* and *Galnt3*$^{-/-}$ mice. To date, several

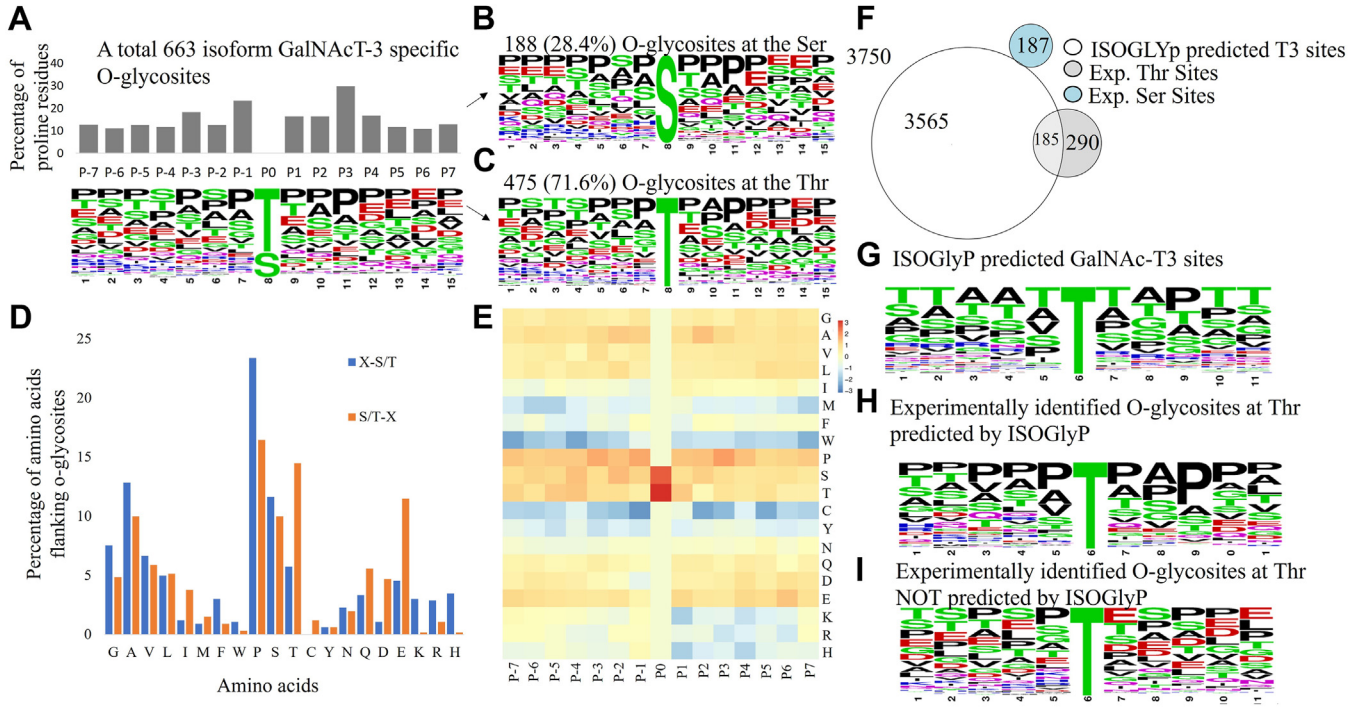


Figure 8. Sequence motif analysis of GalNAc-T3 substrate O-glycosites. *A*, predominant amino acids experimentally identified 663 GalNAc-T3-specific O-glycosites, (*B*) 471 O-glycosites at threonine residues and (*C*) 188 at serine residues. *D*, frequency of amino acids flanking O-glycosite at the −1 and one position; X-S/T and S/T-X,(*E*) Heatmap of the frequency of amino acid residues at ±7 position flanking O-glycosites, the *x*-axis represents the ±7 amino acid position flanking O-glycosites (P0 position); *y*-axis represents amino acids (*F*) Euler diagram showing O-glycosites predicted by ISOGlyP and experimentally identified O-glycosites at threonine and serine residues. *G*, the sequence motif of ISOGlyP predicted 3750 GalNAc-T3 sites (EV > 1). *H*, sequence motif of experimentally identified 190 O-glycosites at threonine predicted by ISOGlyP. *I*, sequence motif of experimentally identified 290 O-glycosites at threonine NOT predicted by ISOGlyP. EV, enhancement value.

approaches either alone or in combination have been used to identify GalNAc-T-specific substrates (28–35). However, none of the mentioned approaches and studies provided information about O-glycosylation sites on proteins in their native, *in vivo* state. Recently, we identified 82 O-glycosites specific to GalNAc-T2 in the liver tissue by comparing substrates from *WT* and *Galnt2-null* mice to gain insights into the precise role of site-specific O-glycosylation in regulating the intricate mechanisms governing metabolic homeostasis (11). Here, we provide an extensive analysis of *in vivo* O-glycosylation sites specific to GalNAc-T3 across multiple tissues in mice.

We used STRING network analysis to gain insights into how the identified GalNAc-T3 substrates are connected to and participate in different biological pathways. Among several other pathways, we detected 55 substrates related to skeletal characteristics, 16 to bone development, and 11 to bone remodeling, thus highlighting the link between GalNAc-T3 and aspects of bone health. However, the specific mechanisms through which they contribute to an abnormal skeletal phenotype can be complex and may involve various pathways and direct as well as indirect interactions. In addition, we identified GalNAc-T3 substrates sites in 14 glycoproteins that are involved in the mineral metabolism. Among them, we identified the store-operated calcium entry regulator STIMATE, which acts as a regulator of store-operated Ca2+ entry at the endoplasmic reticulum-plasma membrane junction (36), as a GalNAc-T3 substrate. While phosphate plays a direct role in the development of HFTC, calcium imbalances may influence the ability to regulate phosphate levels, potentially exacerbating the development of soft tissue calcification. Additionally, we identified GalNAc-T3 specific sites in 49 glycoproteins that bind with calcium. Interestingly, we identified extracellular serine/threonine protein kinase FAM20C that phosphorylates and regulates proteins involved in the biomineralization bone (37). FAM20C is responsible for the phosphorylation of FGF23 at Ser180 and this event inhibits O-glycosylation of FGF23 at T178 by GalNAc-T3, thus, the balance between phosphorylation and O-glycosylation influences the stability and function of FGF23 (38). The alpha-2-HS glycoprotein (Ahsg) T267 site had the highest peptide-spectrum match (PSM) in both tissues and SGs. It was identified as a GalNAc-T3 substrate in muscle with bone and SLGs from both sexes. Although the biological role of Ahsg T267 in HFTC is unclear, we speculate that deglycosylated Ahsg is unable to handle excess minerals formed due to perturbations in calcium phosphate homeostasis and increases the risk of soft tissue calcification. Several enzyme clusters have been identified to play crucial roles in various glycosylation processes. These include St6galnac1, B4galt1, St3gal4, Man1b1, Mgat2, Man2a1, B3gnt6, Man2a2, Prkcsh, Mgat4a, Mgat1, and Pomgnt1 which are involved in O-linked glycosylation, N-glycan processing, glycosphingolipid biosynthesis, and mannose type O-glycan biosynthesis. Their presence suggests

that many enzymes acquire glycosylation sites, which could contribute to their long-term stability, prevent protein aggregation, and ensure correct subcellular localization (39, 40).

Galnt3 is highly expressed in murine SMGs. Therefore, we reasoned to perform in-depth GalNAc-T3 substrate analysis in the glycoprotein and mucin rich SGs from both sexes; and in effect, a total of 488 O-glycosites were identified in the SGs of *Galnt3*$^{-/-}$ mice. GalNAc-T3 is responsible, in part, for the glycosylation of Muc10, the predominant murine salivary mucin. Loss of *Galnt3* results in changes in the structure, composition, and stability of the oral microbiome, suggesting that the loss of *Galnt3* could result in oral disease (9). The glycans on salivary mucins act as recognition motifs and binding sites for certain commensal and pathogenic strains of bacteria in the oral cavity (41). Major salivary glycoproteins such as the salivary mucin MG2, encoded by *MUC7*, the human ortholog of Muc10, serve as prominent targets for bacterial binding suggesting the role of mucins in maintaining the stability of oral microbiome (42). We speculate that variations in the densely O-glycosylated sites in the PTS repeat domain (rich in aminoacids proline, threonine and serine) of mucins might influence how microbes interact with them, including microbial colonization and clearance through glycan-mediated binding.

A significant number of GalNAc-T3-specific sites were identified in critical ECM components such as laminins, collagens, and fibulins as well as cell-cell adhesion molecules such as neurexin-1; 2, podoplanin, cell adhesion molecule 1. These glycoproteins are essential for maintaining tissue integrity and bone structure. Some of the identified O-glycosylation sites on mouse proteins were aligned with their human counterparts (Table S3 and Fig. S3). Altered O-glycosylation of ECM proteins might affect structural integrity, cellular recognition, and interactions with signaling molecules, potentially influencing the microenvironment where ectopic calcifications occur in HFTC. The O-GalNAc sites, S57 on osteocalcin (Bglap) and T64 on osteocalcin related protein (Bglap3), on non-collagenous protein of bone, were identified in the SMGs. The site S57 on recombinant mouse Bglap was verified *in vitro* using murine GalNAc-T3 and GalNAc (13C6) followed by IMPa cleavage (Fig. 5, *D* and *E*). Al Rifai *et al.* reported increased half-life in plasma ex-vivo and circulation *in vivo* in mice due to O-glycosylation at S57 (43). Our future studies will focus on the functional analysis of GalNAc-T3 specific O-glycosites to understand if the loss of these sites influences the complex phenotypes associated.

In our experimental findings, we noticed a strong tendency for GalNAc-T3 sites to have charged acidic amino acids in close proximity to the glycosites. Substrates of GalNAc-T3 exhibited an enhanced presence of hydrophobic residues at positions +1 and + 3 and a high frequency of alanine at the +2 position near O-glycosites (44). The mFGF23 and hFGF23 proteins have negatively charged amino acids, specifically Glu-Asp-Pro and Glu-Asp-Asp respectively, flanking the T178 O-glycosylation site beyond the +3 position at the C-terminal, suggesting a significant role in GalNAc-T3 activity. The charged residues flanking beyond the ± three positions of experimentally identified O-glycosylation indicates that GalNAc-T3 exhibits the highest activity toward substrates with a net neutral charge, followed by those with a net negative charge, and finally those with a net positive charge, However, Ballard *et al.*, report that GalNAc-T3 prefers the most positively charged substrates (45). The exact cause of this discrepancy remains unclear. Interestingly, a significant portion (44%) of threonine sites were not predicted by the ISOGlyP software (https://isoglyp.utep.edu//) we used. It is not clear whether this is due to the training data used to develop the software lacking charged flanking residues, enhanced identification of sequences with positive charges by the mass spectrometer, or if it reflects the actual *in vivo* preference of GalNAc-T3 for acidic residues near O-glycosites.

Our study has certain limitations, one of which is the inability to detect the GalNAc-T3-specific O-glycosite T178 on the FGF23 protein. This glycosite plays a crucial role in inhibiting furin cleavage and facilitating the secretion of fully active FGF23. The flanking region surrounding T178 (^{173}RPRRHT^{178}RS^{180}) contains multiple arginine (R) residues that serve as a digestion site for trypsin. Consequently, during the sample preparation procedure, which includes C18 solid-phase extraction steps, it is possible that very small fragments containing the T178 site may be washed away. Additionally, we were only able to accurately map glycopeptides with GalNAc or Gal-GalNAc structures on the glycopeptides due to the specificity of OgpA enzyme. Most O-glycosites in densely glycosylated mucins were not detected in tissues such as the SGs, colon and lung, presumably due to the inaccessibility of the protein core to trypsin and ambiguities with sequences containing repetitive serine/threonine residues that result in glycopeptides that are too long or too short or the LC-MS analysis after OgpA cleavage. We are developing complementary approaches to address these more complex structures.

We have presented a comprehensive survey of GalNAc-T3-specific O-glycosites spanning multiple tissues at the organism level. The resulting map of GalNAc-T3-specific O-glycosites serves as a valuable resource, representing an important advancement in our efforts to understand the functional implications of O-glycosylation and the phenotypic complexities of patients suffering from HFTC.

Experimental procedures

Animal breeding and genotyping

Dr Michael J. Econs kindly provided the *Galnt3*$^{-/-}$ mice (5). *Galnt3*$^{-/-}$ mice were backcrossed into the C57BL/6MHsd inbred mouse background for at least six generations before analysis. Heterozygous *Galnt3* animals (*Galnt3*$^{+/-}$) were crossed to produce WT and homozygous *Galnt3*-null (*Galnt3*$^{-/-}$) siblings used for all experiments. Genotyping was performed as described previously (9). Experimental procedures were reviewed and approved by the Animal Care and Use Committee of the National Institutes of Health (ASP #19–1016). One-year-old WT and *Galnt3*$^{-/-}$ female mice were the source of all tissues studied, except for the SGs. Two WT and *Galnt3*$^{-/-}$ male and female mice were used as the source

of the SMG and SLG at the age of 6 months. All dissected organs were stored immediately at −80 °C until analysis.

Enrichment and extraction of O-glycopeptides

O-GalNAc glycopeptides were enriched and extracted from tissues using the extraction of O-linked glycopeptides (EXoO) method as described previously (12, 13). Briefly, tissues were dissolved, and proteins were denatured in 8 M urea/200 mM Tris–HCl pH 8.0 using sonication. Proteins were reduced using 5 mM DTT (Thermo Fisher Pierce) at 37 °C for 1 h, alkylated using 10 mM iodoacetamide (Sigma-Aldrich) at room temperature (RT) for 40 min in the dark. The samples were diluted eight times using 200 mM Tris–HCl pH 8.0 and digested using trypsin (Promega) with an enzyme:protein ratio of 1:40 wt/wt to generate peptides. Next, peptides were guanidinated and desalted on the same Sep-Pak C18 cartridge (Waters). For the global proteome analysis, an aliquot of peptides (100 μg) from respective tissues was frozen at −80 °C and dried by vacuum centrifugation. Peptides eluted from C18 in 50% acetonitrile/0.1% trifluoroacetic acid (TFA) were diluted with acetonitrile and TFA to become 95% acetonitrile/1% TFA before glycopeptide enrichment using a HyperSep Retain AX cartridge (Thermo Fisher Scientific). Then, hydrophilic enrichment of the glycopeptides was done using the HyperSep Retain AX (RAX) cartridge (Thermo Fisher Scientific). Glycopeptides were conjugated to a solid phase AminoLink Plus coupling resin (Thermo Fisher Scientific) to remove impurities and contaminated peptides. Recombinant OgpA and sialidase were expressed and purified as described previously (11). Subsequently, O-GalNAc glycopeptides were released using OgpA, and sialic acids were removed using sialidase, using both enzymes at an enzyme:glycopeptide ratio of 1:10 wt/wt. The released O-GalNAc glycopeptides were desalted and eluted with 35% acetonitrile/0.1% TFA, dried using a vacuum concentrator and resuspended in 0.1% TFA.

TMTpro16-plex labeling and fractionation of O-GalNAc glycopeptides and peptides

The peptides and O-glycopeptides were labeled with TMTpro 16plex (Thermo Fisher Scientific) according to manufacturer instructions, allowing for simultaneous identification and quantification of peptides and O-glycopeptides (Fig. 1 for the experimental layout). In brief, glycopeptides (12 μg) and peptides (25 μg) were reconstituted in 100 mM Hepes buffer (pH 8.0) and labeled with TMTpro16-plex reagents in anhydrous acetonitrile at RT for 60 min. Three microliters of each sample were pooled, desalted using C18, and analyzed on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific) to check labeling efficiency. After verification of labeling efficiency, samples were quenched by adding 5% hydroxylamine and pooled. The glycopeptides and pooled peptides were then desalted using 100 mg Sep-Pak solid-phase extraction columns and eluted with 50% acetonitrile/0.1% TFA. Next, labeled samples were fractionated with basic-pH reversed-phase chromatography using a Zorbax Extend-C18 column 600 Bar (1.8 μm, 4.6 × 100 mm; Agilent Technologies) using a 1260 Infinity II HPLC System (Agilent Technologies). Mobile phase A was ammonium formate (10 mM, pH 10), and mobile phase B was 10 mM ammonium formate and 90% acetonitrile (pH 10). TMTpro 16plex-labeled glycopeptides and peptides were separated using a gradient at flow rate of 0.3 ml/min: 0 to 10% B, 10 min; 10 to 20% B, 5 min; 20 to 38% B, 80 min; 38 to 60% B, 5 min; 60 to 95% B, 5 min; and 95-95% B, 15 min. Fractions were collected from 8 to 104 min, and 96 fractions were combined into 24 fractions. The samples were then acidified using TFA and dried with a vacuum concentrator.

Mass spectrometry

TMTpro16-plex-labeled O-glycopeptides and peptide samples were analyzed on an orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific) interfaced to an UltiMate3000 RSLCnano HPLC system (Thermo Fisher Scientific) using a product-dependent method constructed using EThcD triggered scans (HCD-pd-EThcD). One μg of fractionated O-glycopeptides and peptides were injected (5 μl) into an Acclaim PepMap C18 Nano trap column (3 μm, 100 Å, 75 μm × 2 cm) and separated with an Acclaim PepMap C18 Nano column (3 μm, 100 Å, 75 μm × 25 cm). The mobile phases, 0.1% formic acid (Buffer A) and 80% acetonitrile, 0.1% formic acid (Buffer B) were used for nLC. All samples, dissolved in 0.1% TFA, were injected on to the column, held at 0.3 ml/min and eluted in a gradient from 10% to 50% buffer B in 72 min, an increase from 50 to 95% from 72 to 77 min, followed by isocratic flow at 95% buffer B from 77 to 83 min and a reequilibration at 0% of buffer B for 7 min for a total elution time of 90 min.

The mass spectrometer was set to acquire full scan MS spectra (250–1800 m/z) for a maximum injection time of 50 ms at a mass resolution of 120K, RF lens at 30% and an automated gain control target value of 40,000. The precursors of z = 2 to 8 were selected for data-dependent tandem mass spectrometry (MS/MS) scans for 1 s of cycle time, and dynamic exclusion was set to 60 s at an exclusion window of 10 ppm. An isolation window of 0.7 m/z was used to select precursor ions with the quadrupole. HCD-pd-EThcD method was used, where presence of at least two out of nine oxonium ions (126.055, 138.0549, 144.0655, 168.0654, 186.076, 204.0865, 274.0921, 292.1027, and 366.1395) in a higher-energy collisional dissociation (HCD) MS/MS scan triggered acquisition of a second MS/MS scan. Requiring at least two oxonium ions can help improve the specificity of triggering on glycopeptide precursor ions. In the HCD scans, the automated gain control target value was set to 100,000, the collision energy was 38%, the maximum injection time was 120 ms, and a resolution was 50,000. If at least two of the nine listed oxonium ions were present in the scout HCD scan within a ±15 ppm tolerance and were among the 20 most intense peaks, a second MS/MS scan was triggered. In the EThcD method, for charge states 2, 3, 4, and 5 to 10, the

electron-transfer dissociation reaction was set to 125, 100, 75, and 75 ms, respectively, and supplemental activation collision energy of 35% was set at the resolution of orbitrap at 30K.

Data processing for the identification and quantification of O-glycopeptides and peptides

To characterize O-glycopeptides and site-specific O-GalNAc (HexNAc), raw files were searched using software packages-MSFragger-Glyco (Version 17.0, https://msfragger.nesvilab.org/), pGlyco3.0 (https://github.com/pFindStudio/pGlyco3), and O-pair Search with MetaMorpheus (https://github.com/smith-chem-wisc/MetaMorpheus). The searches using MSFragger-Glyco and MetaMorpheus were performed on the Biowulf High-Performance Computing Cluster at NIH (http://hpc.nih.gov). The Uniprot *Mus musculus* database (downloaded on November 12, 2021) was used in all search engines. The following search parameters were used: enzyme: trypsin and OpeRATOR; static modifications: TMTpro (any N terminus), carbamidomethyl (C); dynamic modification: oxidation (M); guanidinyl (K); TMTpro (K), acetylation: protein N terminus. The O-glycan database used for O-glycopeptide searches consisted of common O glycans in search engines (Dataset S13). For MSFragger-Glyco, a basic workflow for glyco-O-HCD was used. For pGlyco3, the HCD + EThcD search mode was selected. The digestion was specific at the C-termini of lysine (K) and arginine (R) residues, and at the N-termini of serine (S) and threonine (T) residues, allowing up to two missed cleavages. Glycopeptide FDR was 0.01, and Percolator was used. For O-pair, the "Glyco search" option was selected. In data type, dissociation type was set as HCD, child scan dissociation was set as EThcD, and search parameters were set as described previously (16). Further data analysis was done in Excel. Output data were processed by filtering PSMs containing HexNAc in the observed modifications, peptides starting with serine or threonine residues, having glycan q-value less than 0.01. The result was combined based on the file names and MS2 scan numbers. PSMs with the EThcD O-glycosite mapping score of at least 0.75 on the first serine or threonine residues of peptide sequences were used for quantitation analysis. Proteome Discoverer 2.5 (PD 2.5) (Thermo Fisher Scientific) was used to extract TMTpro reporter ion signal-to-noise (S/N) from the raw files and then the S/N numbers were matched to the assigned PSMs according to the file names and MS2 scan number. Data were normalized using the median of each TMT channel.

TMTpro16 plex-labeled proteomics data were processed using PD2.5. MS spectra were searched against the Uniport *M. musculus* database (downloaded on November 12, 2021) and contaminant database using SEQUEST HT. Percolator Target FDR (Strict) was set to 0.01; percolator relaxed FDR:0.05; and protein filter: minimum number of peptide sequences: 2. For quantification, the Reporter ion quantifier node was applied for TMTpro16 plex quantification of reporter ions using total peptide amount normalization and spectra with more than 50% isolation interference were excluded. Protein quantification was performed using unique and razor peptides, and protein ratio calculation was based on protein abundance.

Data were merged from these three software packages, together with reporter ion intensities, and global proteomics. This allowed us to identify O-glycopeptides, their corresponding glycans, the site-specific localization of glycosylation, quantification of O-glycopeptides, and protein level change. The relative abundance of O-glycopeptides and peptides was determined by calculating the ratio of median normalized intensity of *Galnt3*$^{-/-}$ *versus* WT mice. The 1.5 cutoff was determined based on the observation that most glycopeptide intensity ratios from WT technical replicates within the dataset ranged from 0.65 to 1.5, suggesting that values outside this range indicate significant differences. A reduced fold change of at least 1.5 in the intensity of O-glycopeptides and no protein level change between *Galnt3*$^{-/-}$ and WT samples were used to define GalNAc-T3-specific O-glycosites.

Validation of O-GalNAc glycosites

Recombinant mouse osteocalcin expressed in *Escherichia coli* was purchased from Abcam. The glycoprotein was reduced using 5 mM DTT (Thermo Fisher Pierce) at 37 °C for 1 h, alkylated using 10 mM iodoacetamide (Sigma-Aldrich) at RT for 40 min in the dark and digested using trypsin (Promega, enzyme:protein ratio of 1:40 wt/wt). Peptides were desalted by C18, dried using a vacuum concentrator, and resuspended in 50 mM Tris–HCl buffer pH 8.0. Murine GalNAc-T3 was expressed using COS7 cells as described previously (46). Glycosylation reactions were performed as described previously (46) with modification, using 10 µl of media from COS7 cells transfected with murine Galnt3, 5 mM UDP-GalNAc($^{13}C_6$) (Chemily Glycoscience), 5 mM $MnCl_2$ in reaction buffer 25 mM Tris–HCl, pH 7.4. Reactions were run at 37 °C for 16 h and quenched with 1% TFA. Reactions were subjected to C18 desalting, drying in a vacuum concentrator, and resuspended in 50 mM Tris-HCl buffer pH 8.0. Recombinant immunomodulating metalloprotease (IMPa) was expressed and purified as described previously (11). IMPa was added to cleave the N termini of serine and threonine residues glycosylated by GalNAc($^{13}C_6$). Finally, the peptides were desalted by C18 and dried using a vacuum concentrator. Trans-Golgi network integral membrane protein 1/2 (Tgoln1/2) was purified from male SLGs derived from *WT* and *Galnt3*$^{-/-}$ mice to validate a subset of O-glycosites identified as described in SI Appendix. *In vitro* enzyme activity on peptides was performed to verify GalNAc-T3 selectively along with other isoforms expressed in the representative tissue, SGs, as described in SI Appendix.

Data visualization

WebLogo was used for consensus motif analysis (https://weblogo.berkeley.edu/logo. cgi) (47). DAVID: a web server tool was used for the functional enrichment analysis of O-glycoproteins that corresponded to identified O-glycosite substrates (48). Heatmaps and plots were generated using ImageGP: a data visualization web server (49) and SRtools (50). Venn diagram graphs were generated using InteractiVenn (51). Euler diagrams were generated using https://eulerr.co/. Protein-protein network analysis was generated using STRING version 12.0 (52). Isoform Specific O-Glycosylation Prediction

(ISOGlyP, https://isoglyp.utep.edu/) was used to predict GalNAc-T isoenzyme-specific glycopeptide substrates. Volcano plots and Dot plots were generated using R package.

Data availability

The mass spectrometry raw files been deposited to the ProteomeXchange Consortium *via* the PRIDE (53) partner repository with the dataset identifier PXD047517. The raw data files related to validation experiments can be accessed using the identifier PXD052978 and PXD053014. All other data are included in the manuscript and/or supporting information.

Supporting information—This article contains supporting information (10, 54, 55).

Acknowledgments—This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov). We thank our colleagues for helpful discussions.

Author contributions—K. D., W. Y., K. G. T. H., and L. A. T. visualization; K. D., A. C., P. M., and A. J. L. investigation; K. D. and W. Y., formal analysis; K. D. and E. T. methodology; K. D. and L. A. T. conceptualization; K. D. writing–original draft; W. Y., K. G. T. H., and L. A. T. writing–review and editing; E. T. resources; K. G. T. H. and L. A. T. supervision; K. G. T. H. and L. A. T. funding acquisition.

Funding and additional information—This research was supported by the Intramural Research Program of the NIDCR, NIH (Z01-DE-000713 to K. G. T. H. and 1-ZIA707 DE000739–05 to L. A. T.) and NIDCR Mass Spectrometry Facility (ZIA DE000751). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CDG, congenital disorders of glycosylation; ECM, extracellular matrix; EV, enhancement value; FDR, false discovery rate; FGF23, fibroblast growth factor 23; Galnt, polypeptide N-acetylgalactosaminyltransferase; HCD, higher-energy collisional dissociation; HFTC, hyperphosphatemic familial tumoral calcinosis; MS/MS, tandem mass spectrometry; PSM, peptide-spectrum match; RT, room temperature; SGs, salivary glands; SLGs, sublingual glands; SMGs, submandibular glands; TMT, tandem mass tag; TFA, trifluoroacetic acid.

References

1. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736–756
2. Tran, D. T., and Ten Hagen, K. G. (2013) Mucin-type O-glycosylation during development. *J. Biol. Chem.* **288**, 6921–6929
3. Freeze, H. H. (2006) Genetic defects in the human glycome. *Nat. Rev. Genet.* **7**, 537–551
4. Topaz, O., Shurman, D. L., Bergman, R., Indelman, M., Ratajczak, P., Mizrachi, M., *et al.* (2004) Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nat. Genet.* **36**, 579–581
5. Ichikawa, S., Sorenson, A. H., Austin, A. M., Mackenzie, D. S., Fritz, T. A., Moh, A., *et al.* (2009) Ablation of the Galnt3 gene leads to low-circulating intact fibroblast growth factor 23 (Fgf23) concentrations and hyperphosphatemia despite increased Fgf23 expression. *Endocrinology* **150**, 2543–2550
6. Dumitrescu, C. E., Kelly, M. H., Khosravi, A., Hart, T. C., Brahim, J., White, K. E., *et al.* (2009) A case of familial tumoral calcinosis/hyperostosis-hyperphosphatemia syndrome due to a compound heterozygous mutation in GALNT3 demonstrating new phenotypic features. *Osteoporos. Int.* **20**, 1273–1278
7. Ichikawa, S., Gray, A. K., Padgett, L. R., Reilly, A. M., and Unsicker, T. R. (2014) High dietary phosphate intake induces development of ectopic calcifications in a murine model of familial tumoral calcinosis. *J. Bone Miner Res.* **29**, 2017–2023
8. Orr, S. L., Le, D., Long, J. M., Sobieszczuk, P., Ma, B., Tian, H., *et al.* (2013) A phenotype survey of 36 mutant mouse strains with gene-targeted defects in glycosyltransferases or glycan-binding proteins. *Glycobiology* **23**, 363–380
9. Peluso, G., Tian, E., Abusleme, L., Munemasa, T., Mukaibo, T., and Ten Hagen, K. G. (2020) Loss of the disease-associated glycosyltransferase Galnt3 alters Muc10 glycosylation and the composition of the oral microbiome. *J. Biol. Chem.* **295**, 1411–1425
10. Bennett, E. P., Hassan, H., and Clausen, H. (1996) cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D-galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3. *J. Biol. Chem.* **271**, 17006–17012
11. Yang, W., Tian, E., Chernish, A., McCluggage, P., Dalal, K., Lara, A., *et al.* (2023) Quantitative mapping of the in vivo O-GalNAc glycoproteome in mouse tissues identifies GalNAc-T2 O-glycosites in metabolic disorder. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2303703120
12. Yang, W., Song, A., Ao, M., Xu, Y., and Zhang, H. (2020) Large-scale site-specific mapping of the O-GalNAc glycoproteome. *Nat. Protoc.* **15**, 2589–2610
13. Yang, W., Ao, M., Hu, Y., Li, Q. K., and Zhang, H. (2018) Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). *Mol. Syst. Biol.* **14**, e8486
14. Polasky, D. A., Yu, F., Teo, G. C., and Nesvizhskii, A. I. (2020) Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **17**, 1125–1132
15. Zeng, W. F., Cao, W. Q., Liu, M. Q., He, S. M., and Yang, P. Y. (2021) Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat. Methods* **18**, 1515–1523
16. Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R., and Smith, L. M. (2020) O-pair search with MetaMorpheus for O-glycopeptide characterization. *Nat. Methods* **17**, 1133–1138
17. Culp, D. J., Robinson, B., Cash, M. N., Bhattacharyya, I., Stewart, C., and Cuadra-Saenz, G. (2015) Salivary mucin 19 glycoproteins: innate immune functions in Streptococcus mutans-induced caries in mice and evidence for expression in human saliva. *J. Biol. Chem.* **290**, 2993–3008
18. Kameyama, A., Yamakoshi, K., and Watanabe, A. (2019) A rapid separation and characterization of mucins from mouse submandibular glands by supported molecular matrix electrophoresis. *Biochim. Biophys. Acta Proteins Proteom* **1867**, 76–81
19. Halleen, J. M., Alatalo, S. L., Janckila, A. J., Woitge, H. W., Seibel, M. J., and Vaananen, H. K. (2001) Serum tartrate-resistant acid phosphatase 5b is a specific and sensitive marker of bone resorption. *Clin. Chem.* **47**, 597–600
20. Masayoshi, Y. (2000) Role of regucalcin in calcium signaling. *Life Sci.* **66**, 1769–1780
21. Huang, J., Wu, M., Zhang, Y., Kong, S., Liu, M., Jiang, B., *et al.* (2021) OGP: a repository of experimentally characterized O-glycoproteins to facilitate studies on O-glycosylation. *Genomics Proteomics Bioinform.* **19**, 611–618
22. Downs, M., Curran, J., Zaia, J., and Sethi, M. K. (2023) Analysis of complex proteoglycans using serial proteolysis and EThcD provides deep

N- and O-glycoproteomic coverage. *Anal. Bioanal. Chem.* **415**, 6995–7009
23. Klein, J. A., Meng, L., and Zaia, J. (2018) Deep sequencing of complex proteoglycans: a novel strategy for high coverage and site-specific identification of glycosaminoglycan-linked peptides. *Mol. Cell Proteomics* **17**, 1578–1590
24. Lee, N. K., Sowa, H., Hinoi, E., Ferron, M., Ahn, J. D., Confavreux, C., *et al.* (2007) Endocrine regulation of energy metabolism by the skeleton. *Cell* **130**, 456–469
25. Yue, R., Shen, B., and Morrison, S. J. (2016) Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton. *Elife* **5**, e18782
26. Bassett, J. H., Gogakos, A., White, J. K., Evans, H., Jacques, R. M., van der Spek, A. H., *et al.* (2012) Rapid-throughput skeletal phenotyping of 100 knockout mice identifies 9 new genes that determine bone strength. *PLoS Genet.* **8**, e1002858
27. Zeng, F., and Harris, R. C. (2014) Epidermal growth factor, from gene organization to bedside. *Semin. Cell Dev. Biol.* **28**, 2–11
28. Kong, Y., Joshi, H. J., Schjoldager, K. T., Madsen, T. D., Gerken, T. A., Vester-Christensen, M. B., *et al.* (2015) Probing polypeptide GalNAc-transferase isoform substrate specificities by in vitro analysis. *Glycobiology* **25**, 55–65
29. Gerken, T. A., Raman, J., Fritz, T. A., and Jamison, O. (2006) Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *J. Biol. Chem.* **281**, 32403–32416
30. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., *et al.* (2015) Deconstruction of O-glycosylation–GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep.* **16**, 1713–1722
31. Li, X., Zhang, Y., Zhang, M., and Wang, Y. (2020) GALNT2 regulates ANGPTL3 cleavage in cells and in vivo of mice. *Sci. Rep.* **10**, 16168
32. Hintze, J., Ye, Z., Narimatsu, Y., Madsen, T. D., Joshi, H. J., Goth, C. K., *et al.* (2018) Probing the contribution of individual polypeptide GalNAc-transferase isoforms to the O-glycoproteome by inducible expression in isogenic cell lines. *J. Biol. Chem.* **293**, 19064–19077
33. Narimatsu, Y., Joshi, H. J., Schjoldager, K. T., Hintze, J., Halim, A., Steentoft, C., *et al.* (2019) Exploring regulation of protein O-glycosylation in isogenic human HEK293 cells by differential O-glycoproteomics. *Mol. Cell Proteomics* **18**, 1396–1409
34. Bagdonaite, I., Pallesen, E. M., Ye, Z., Vakhrushev, S. Y., Marinova, I. N., Nielsen, M. I., *et al.* (2020) O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. *EMBO Rep.* **21**, e48885
35. Nielsen, M. I., de Haan, N., Kightlinger, W., Ye, Z., Dabelsteen, S., Li, M., *et al.* (2022) Global mapping of GalNAc-T isoform-specificities and O-glycosylation site-occupancy in a tissue-forming human cell line. *Nat. Commun.* **13**, 6257
36. Jing, J., He, L., Sun, A., Quintana, A., Ding, Y., Ma, G., *et al.* (2015) Proteomic mapping of ER-PM junctions identifies STIMATE as a regulator of Ca(2)(+) influx. *Nat. Cell Biol.* **17**, 1339–1347
37. Tagliabracci, V. S., Engel, J. L., Wen, J., Wiley, S. E., Worby, C. A., Kinch, L. N., *et al.* (2012) Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* **336**, 1150–1153
38. Tagliabracci, V. S., Engel, J. L., Wiley, S. E., Xiao, J., Gonzalez, D. J., Nidumanda Appaiah, H., *et al.* (2014) Dynamic regulation of FGF23 by Fam20C phosphorylation, GalNAc-T3 glycosylation, and furin proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5520–5525
39. Ryslava, H., Doubnerova, V., Kavan, D., and Vanek, O. (2013) Effect of posttranslational modifications on enzyme function and assembly. *J. Proteomics* **92**, 80–109
40. Mikolajczyk, K., Kaczmarek, R., and Czerwinski, M. (2020) How glycosylation affects glycosylation: the role of N-glycans in glycosyltransferase activity. *Glycobiology* **30**, 941–969
41. Cross, B. W., and Ruhl, S. (2018) Glycan recognition at the saliva - oral microbiome interface. *Cell Immunol.* **333**, 19–33
42. Ruhl, S., Sandberg, A. L., and Cisar, J. O. (2004) Salivary receptors for the proline-rich protein-binding and lectin-like adhesins of oral actinomyces and streptococci. *J. Dent. Res.* **83**, 505–510
43. Al Rifai, O., Julien, C., Lacombe, J., Faubert, D., Lira-Navarrete, E., Narimatsu, Y., *et al.* (2020) The half-life of the bone-derived hormone osteocalcin is regulated through O-glycosylation in mice, but not in humans. *Elife* **9**, e61174
44. Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., *et al.* (2011) Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* **286**, 14493–14507
45. Ballard, C. J., Paserba, M. R., Paul Daniel, E. J., Hurtado-Guerrero, R., and Gerken, T. A. (2023) Polypeptide N-acetylgalactosaminyltransferase (GalNAc-T) isozyme surface charge governs charge substrate preferences to modulate mucin type O-glycosylation. *Glycobiology* **33**, 817–836
46. Zara, J., Hagen, F. K., Ten Hagen, K. G., Van Wuyckhuyse, B. C., and Tabak, L. A. (1996) Cloning and expression of mouse UDP-GalNAc: PolypeptideN-acetylgalactosaminyltransferase-T3. *Biochem. Biophysical Res. Commun.* **228**, 38–44
47. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
48. Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., *et al.* (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221
49. Chen, T., Liu, Y. X., and Huang, L. (2022) ImageGP: an easy-to-use data visualization web server for scientific researchers. *iMeta* **1**, e5
50. Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., *et al.* (2023) SRplot: a free online platform for data visualization and graphing. *PLoS One* **18**, e0294236
51. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., and Minghim, R. (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169
52. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., *et al.* (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646
53. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552
54. de Las Rivas, M., Daniel, P., Narimatsu, Y., Companon, I., Kato, K., Hermosilla, P., *et al.* (2020) Molecular basis for fibroblast growth factor 23 O-glycosylation by GalNAc-T3. *Nat. Chem. Biol.* **16**, 351–360
55. Nehrke, K., Hagen, F. K., and Tabak, L. A. (1998) Isoform-specific O-glycosylation by murine UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-T3, in vivo. *Glycobiology* **8**, 367–371