GOLPH3 and GOLPH3L are broad-spectrum COPI adaptors for sorting into intra-Golgi transport vesicles

Lawrence G. Welch[1,2], Sew-Yeu Peak-Chew[1], Farida Begum[1], Tim J. Stevens[1] and Sean Munro[1,2]

1: MRC Laboratory of Molecular Biology
Francis Crick Avenue
Cambridge CB2 0QH
UK

2: Correspondence: lwelch@mrc-lmb.cam.ac.uk; sean@mrc-lmb.cam.ac.uk

Abstract

Glycosylation is a diverse and abundant modification of proteins, lipids and RNA. The fidelity of glycosylation is, in part, assured by the correct compartmentalisation of Golgi-resident glycosylation enzymes within the Golgi stack. The COPI adaptor GOLPH3 has been shown to interact with the cytoplasmic tails of a subset of Golgi enzymes and direct their retention in the Golgi. However, other mechanisms of retention, and other roles for GOLPH3, have been proposed, and a comprehensive characterisation of the clientele of GOLPH3 and its paralogue GOLPH3L has been lacking. The role of GOLPH3 is of particular interest as it is frequently amplified in several solid tumour types. Here, we combine two orthogonal proteomic analyses to identify a diverse range of GOLPH3+3L clients and find that they act in a wide spectrum of glycosylation pathways, or have other roles in the Golgi. Using binding studies, bioinformatics and an *in vivo* Golgi retention assay, we show that GOLPH3+3L interact with the cytoplasmic tails of their clients through membrane-proximal positively-charged residues. Furthermore, deletion of GOLPH3+3L causes diverse defects in glycosylation. Thus, GOLPH3+3L are major COPI adaptors that impinge on most, if not all, of the glycosylation pathways of the Golgi.

Introduction

Glycosylation is one of the most widespread and heterogeneous post-translational modifications that can be attached to a plethora of target substrates including proteins, lipids and RNA (Schjoldager et al., 2020; Maccioni et al., 2011; Sandhoff and Sandhoff, 2018; Flynn et al., 2021). Glycans can have significant impact on the structure, function and stability of biomolecules, and as a result, glycosylation plays an influential role in many pathological and physiological processes (Pinho and Reis, 2015; Tran and Ten Hagen, 2013; Vajaria and Patel, 2017; Pascoal et al., 2020; Stowell et al., 2015).

Secreted proteins, and membrane proteins that traverse or reside in the secretory pathway, are predominantly glycosylated in the endoplasmic reticulum (ER) and Golgi during biogenesis (Moremen et al., 2012). Secretory glycosylation involves the sequential addition of glycan moieties, and this controlled sequence of modification is, in part, dependent on the correct compartmentalisation of specific glycosylation enzymes across the ER and the different cisternae of the Golgi stack (Moremen et al., 2012; Schjoldager et al., 2020). There are approximately a dozen different glycan modification pathways that act on N-linked glycans, O-linked glycans or glycolipids, often with each reaction requiring a unique enzyme (Schjoldager et al., 2020). As a result, the human genome has over 200 genes encoding glycosylation enzymes, many of which are Golgi-resident type II transmembrane proteins (Lombard et al., 2014). These Golgi enzymes typically have a short cytoplasmic N-terminus, a relatively short transmembrane domain (TMD) and an unstructured stem region which acts as a flexible linker between the lipid bilayer and the lumenal catalytic domain (Tu and Banfield, 2010; Welch and Munro, 2019).

For several glycosylation enzymes, the cytoplasmic tail, TMD and stem (known as the CTS domain) have been shown to be responsible for targeting to the correct sub-Golgi location (Tu and Banfield, 2010; Welch and Munro, 2019). It is likely that the CTS domains act by directing the incorporation of the enzymes into budding COPI vesicles which then recycle them within the Golgi stack (Welch and Munro, 2019; Lujan and Campelo, 2021; Adolf et al., 2019). COPI vesicles are generated by the heptameric coatomer complex and auxiliary proteins including the small GTPase Arf1 (Dodonova et al., 2017; Gomez-Navarro and Miller, 2016). According to the cisternal maturation model, Golgi cisternae continually progress from the cis-Golgi to the trans-Golgi, whilst Golgi-residents are segregated away from anterograde cargo into COPI vesicles which bud from the maturing cisternae (Pantazopoulou and Glick, 2019; Glick and Nakano, 2009). These COPI vesicles serve to retrieve Golgi-resident cargoes and deliver them to their correct cisternal location, against the flow of the maturing cisternae. Although other models for Golgi organisation have been proposed, the incorporation of Golgi enzymes into intra-Golgi COPI vesicles is increasingly well established (Adolf et al., 2019; Dunlop et al., 2017). However, what is less well understood are the mechanisms by which the CTS domains of the many different enzymes direct them into budding COPI vesicles. Moreover, it is unclear if and how these mechanisms differ between vesicles budding from different parts of the Golgi stack, especially as vesicles for retrograde traffic from the early Golgi to the ER are also formed

by the COPI coat. Golgi enzymes vary in their distribution across the Golgi stack, implying that there are distinct sorting signals in their CTS domains which serve to maintain this heterogenous distribution and thus ensure the fidelity of glycosylation (Lujan and Campelo, 2021; Welch and Munro, 2019; Kornfeld and Kornfeld, 1985).

The membrane thickness model proposes that Golgi-residents with relatively short TMDs favour a thinner bilayer in budding COPI vesicles over a thick, sphingolipid/sterol-rich membrane that is formed at the late-Golgi and then proceeds to post-Golgi compartments (Bretscher and Munro, 1993; Sharpe et al., 2010; van Galen et al., 2014). Other, complimentary, models propose that the cytoplasmic tails of Golgi residents interact with the COPI coat, either directly or indirectly through COPI adaptors. Several cis-Golgi-resident enzymes have been reported to bind directly to the COPI coat through a φ(K/R)XLX(K/R) motif in their cytoplasmic tails (Liu et al., 2018). In addition, the COPI adaptor GOLPH3 and its yeast ortholog Vps74 have been shown to be required for the Golgi retention of a selection of glycosyltransferases, and in some cases found to bind directly to their cytoplasmic tails (Tu et al., 2008; Schmitz et al., 2008; Isaji et al., 2014; Ali et al., 2012; Chang et al., 2013; Pereira et al., 2014). GOLPH3/Vps74 has been proposed to be recruited to the trans-Golgi through an interaction with PtdIns4P (Dippold et al., 2009; Wood et al., 2009). Once on the membrane, GOLPH3/Vps74 can simultaneously interact with the COPI coat and sample the tails of the enzyme cargo to package them into vesicles recycling from the trans-Golgi to the medial-Golgi (Tu et al., 2008; Schmitz et al., 2008; Eckert et al., 2014). Deletion or depletion of GOLPH3/Vps74 can cause the mislocalisation of its clients to the lysosome or vacuole for degradation (Schmitz et al., 2008; Tu et al., 2008; Rizzo et al., 2021).

Whilst there seems good evidence that GOLPH3 can direct particular enzymes into COPI vesicles, the scale of its contribution to Golgi enzyme retention is still unclear. For instance, it has been recently proposed that GOLPH3 specifically regulates the retention of enzymes involved in glycosphingolipid synthesis (Rizzo et al., 2021). Moreover, several other roles have been proposed for GOLPH3, including regulating Golgi morphology and forward transport from the trans-Golgi network (TGN), raising the possibility that some of the effects on retention may indirect (Rahajeng et al., 2019; Dippold et al., 2009). In addition, roles for GOLPH3 have been evoked in mTOR signalling and the response to DNA damage (Farber-Katz et al., 2014; Scott et al., 2009). Finally, GOLPH3 has been found to be frequently amplified in various solid tumour types, and its overexpression is associated with poor prognosis (Sechi et al., 2015, 2020; Rizzo et al., 2017). Resolution of the role of GOLPH3 could thus benefit from a comprehensive characterisation of its contribution to Golgi enzyme retention. We have therefore applied two orthogonal, non-biased, proteomic analyses to identify clients for GOLPH3, and extended this to GOLPH3L, a paralogue that is expressed at low levels in most tissues but whose function is unclear. By using a combination of *in vitro* binding studies, bioinformatic analyses and an *in vivo* Golgi retention assay, we show that both GOLPH3 and GOLPH3L interact with the short cytoplasmic tails of numerous Golgi residents through membrane-proximal polybasic stretches. Deletion of both GOLPH3 genes triggers an instability in their clientele

which leads to global defects in glycosylation. Thus, GOLPH3 and GOLPH3L are major, broad-spectrum, cargo adaptors for COPI-coated intra-Golgi vesicles.

Results

GOLPH3 and GOLPH3L bind a diverse array of Golgi-resident proteins and the COPI coat

Initially, we used affinity chromatography to identify interactors of GOLPH3 and GOLPH3L. GST fusions to the N-terminus of both GOLPH3 proteins were used for chromatography of HEK293T cell lysate. When compared to GST, both GST-tagged GOLPH3 and GOLPH3L enriched a large number of proteins including many Golgi-resident glycosylation enzymes (referred to by their gene names for simplicity), and all of the subunits of the COPI coat (Fig. 1 A; and Table S1). Immunoblotting confirmed the specific enrichment of GALNT7 and β-COP, a Golgi glycosylation enzyme and a COPI subunit, respectively (Fig. 1 B). Amongst the interacting Golgi enzymes identified by mass spectrometry, several of the previously reported cargo interactors were identified as hits including GCNT1, EXT1, EXT2, GALNT12, POMGNT1, ST3GAL4 and B4GALT5 (Rizzo et al., 2021; Pereira et al., 2014; Eckert et al., 2014; Isaji et al., 2014; Chang et al., 2013; Ali et al., 2012). Comparing the 73 Golgi-resident membrane proteins enriched by either GOLPH3 or GOLPH3L ($P < 0.05$ compared to GST alone), there was a high degree of overlap (42 common hits, 13 GOLPH3-specific and 18 GOLPH3-specific). In total, 692 proteins were enriched by either GOLPH3 or GOLPH3L (Table S2), with a large proportion being proteins from non-Golgi organelles suggesting that there is also considerable non-specific binding. Comparing this GOLPH3+3L interactome to a previously reported COPI proteome generated from HeLa cells revealed that of the 249 proteins of the COPI proteome, 102 proteins (41.0%) were also isolated from cell lysate by GOLPH3+3L (Table S2) (Adolf et al., 2019). Most of these proteins have not been previously reported to bind either GOLPH3 or GOLPH3L, but many are type II Golgi enzymes. The large proportion of the COPI cargo that are GOLPH3+3L interactors suggests that GOLPH3+3L are broad-spectrum adaptors for COPI-coated vesicles.

The tails of GALNT2 and ST6GAL1 are sufficient for GOLPH3+3L-dependent Golgi retention

To validate some of the putative GOLPH3+3L clients identified by affinity chromatography, a selection were examined in vivo. Since GOLPH3 is a cytosolic protein that binds the cytoplasmic tails of type II membrane proteins, we used a reporter based on the type II plasma membrane protein sucrase-isomaltase (SI) fused to GFP (Fig. 2 A, Liu et al., 2018). We then replaced the cytoplasmic tail of the reporter with the cytoplasmic tails of either a novel GOLPH3 client (GALNT2) or a previously reported one (ST6GAL1) (Isaji et al., 2014; Eckert et al., 2014). These reporters were stably integrated into wild-type U2OS cells, or those from which both *GOLPH3* and *GOLPH3L* had been deleted by CRISPR-Cas9 gene-editing (Fig. 2 B; and Fig. S1). As expected, the SI reporter displayed a robust cell surface localisation in both wild-type and Δ*GOLPH3;*Δ*GOLPH3L* U2OS cells (Fig. 2 C). In contrast, the ST6GAL1 and GALNT2 cytoplasmic tail chimeras exhibited a strong Golgi localisation in wild-type cells, but in Δ*GOLPH3;*Δ*GOLPH3L* cells a considerable proportion localised at the plasma membrane in addition to the Golgi. This is consistent with previous reports that ST6GAL1 is a GOLPH3 client, and demonstrates that the affinity chromatography has identified a novel client in GALNT2 (Eckert et al., 2014; Isaji et al., 2014; Liu et al., 2018).

A quantitative Golgi retention assay to interrogate tail- and TMD-dependent retention mechanisms

In order to quantify the phenomenon observed by immunofluorescence, we used a flow cytometry-based assay. The principle of the assay is that GFP-tagged reporters that are retained in the Golgi will not be accessible to an Alexa Fluor 647-conjugated anti-GFP antibody added externally under non-permeabilising conditions (Fig. 3 A). Thus, the ratio of the A647 signal (cell surface signal) to the GFP signal (total cell signal), provides a quantitative measure of retention. As a proof of principle, the GALNT2 cytoplasmic tail chimera and SI reporter cell lines were tested. In wild-type cells, the SI plasma membrane reporter exhibited a linear relationship between cell surface and total cell signals with a high ratio between the two, indicative of efficient exocytosis to the plasma membrane (Fig. 3 B; and Fig. S2). In contrast, the GALNT2 reporter had a low ratio of cell surface to total signal, indicative of Golgi retention. Only at very high levels of expression was the reporter detectable at the surface, indicating saturation of retention. Strikingly, when the GALNT2 reporter was expressed in a *ΔGOLPH3;ΔGOLPH3L* background it behaved like the SI plasma membrane reporter, confirming that Golgi retention was lost upon the deletion of both GOLPH3 genes.

We next applied this quantitative assay to a wider array of reporters. When the SI plasma membrane reporter was expressed in a *ΔGOLPH3;ΔGOLPH3L* background, its behaviour was indistinguishable from that in wild-type cells: the reporter displayed minimal Golgi retention and robust plasma membrane localisation (Fig. 3 C). Although it has been proposed that GOLPH3 is required for efficient anterograde traffic of cargo from the Golgi to the plasma membrane, we could not detect an obvious defect in the traffic of SI upon deletion of *GOLPH3* and *GOLPH3L* (Dippold et al., 2009; Rahajeng et al., 2019). Consistent with the immunofluorescence data, the ST6GAL1 cytoplasmic tail conferred retention in wild-type cells and this was mostly relieved in *ΔGOLPH3;ΔGOLPH3L* cells. We also tested a reporter in which the TMD of SI was replaced with that of ST6GAL1, as its relatively short TMD has previously been shown to be sufficient for Golgi targeting (Munro, 1991; Sun et al., 2021). The ST6GAL1 TMD chimera also exhibited robust Golgi retention but this was independent of GOLPH3/3L, consistent with the model that GOLPH3 proteins specifically recognise the tails, not the TMDs, of their clients. It has also been reported that the tail of GlcNAc-1-phosphotransferase (GNPTAB) can interact directly with the COPI coat, and that the Golgi retention of a GNPTAB cytoplasmic tail chimera is independent of GOLPH3 (Liu et al., 2018). In accordance, with these results, the GNPTAB tail conferred robust Golgi retention which was unperturbed by the deletion of *GOLPH3* and *GOLPH3L* (Fig. 3 D).

A wide range of Golgi-resident proteins are destabilised by the deletion of GOLPH3 genes

Knockdown of GOLPH3 in mammalian cells, or the deletion of its ortholog Vps74 in yeast, has been found to cause the mislocalisation of particular Golgi enzymes to the lysosome or vacuole where they are degraded (Tu et al., 2008; Schmitz et al., 2008; Rizzo et al., 2021; Chang et al., 2013). We therefore tested the effect of removing GOLPH3/3L on the stability of two interactors found by affinity chromatography (GALNT7 and GPP130/GOLIM4), and found that the levels of both were greatly

reduced in the double knockout background (Fig. 4 A). To test the ability of the individual GOLPH3 proteins to rescue this phenotype, each was reintroduced separately using PiggyBac transposition. In both polyclonal populations, the levels of GALNT7 and GOLIM4 were partly restored (Fig. 4 A). Immunofluorescence of the GOLPH3-transduced population revealed considerable heterogeneity in expression levels, suggesting that the partial rescues reflect the presence of low- or non-expressing cells in the polyclonal populations (Fig. S3 A). For GOLPH3 it was possible to clone individual lines that showed uniform expression and rescue, but this was not possible for GOLPH3L suggesting that its over-expression may have a dominant negative effect. Nonetheless, when GOLPH3 or GOLPH3L were transiently transfected into the double-knockout cell line, in both cases cells with a robust rescue of the Golgi accumulation of GALNT7 were clearly present within the population (Fig. S3 B). Thus, the instability of these Golgi residents in the double knockout is a consequence of the loss of the targeted genes, and both GOLPH3 and GOLPH3L can rescue this, and hence confer Golgi retention, individually.

Proteome wide analysis of proteins dependent on GOLPH3/GOLPH3L for their stability

The clear effect of removing GOLPH3 and GOLPH3L on the levels of GALNT7 and GOLIM4, suggested that a global analysis of protein levels could complement the affinity chromatography as an approach to identifying clients of the GOLPH3 proteins. Thus, we used multiplexed quantitative mass spectrometry based on tandem mass tagging (TMT) to compare wild-type, *ΔGOLPH3;ΔGOLPH3L,* and the *GOLPH3*-rescued cells. This revealed that in the double knockout, many Golgi-residents were depleted relative to the wild-type and rescue cell lines (Fig. 4 B; Fig. S1 D; and Table S3). Moreover, additional proteins, including glycoproteins, also showed changes in abundance. This may in part reflect the use of tandem-mass spectrometry which increases the sensitivity of protein detection and hence proteome coverage, but comes at the cost of slightly reduced accuracy of quantitation. In addition, it is also known that changes in glycosylation can affect the stability of glycoproteins (Kingsley et al., 1986; Scott and Panin, 2014; Jayaprakash and Surolia, 2017). Thus, to filter for high confidence GOLPH3+3L clients, we compared the data from the proteomic analysis to that obtained with affinity chromatography and found that that a set of proteins were strongly depleted in the *ΔGOLPH3;ΔGOLPH3L* cell line and also bound efficiently to GST-GOLPH3+3L (Fig. 4 C). Setting a stringent cut-off based on depletion that was greater than the most depleted non-Golgi protein, revealed 22 hits, 17 of which are known Golgi enzymes with the others being Golgi proteins of unknown function, and all but three of the 22 are Type II proteins with a single TMD near the N-terminus (Fig. 4 C; and Table S3). The 17 Golgi enzymes come from a broad array of enzymatic pathways including N-linked glycosylation, O-linked mucin-type glycosylation, proteoglycan synthesis, O-mannosylation, glycosphingolipid synthesis, as well as Golgi enzymes involved in tyrosine sulphation and nucleotide hydrolysis (Fig. 4 D). Just below the strict cut-off used here, were several additional Golgi enzymes, again from a wide range of pathways (Fig. 4 C; and Table S3). Thus, combining the two rather noisy datasets reveals a clear set of proteins that are strong candidates to be clients for GOLPH3/3L-dependent Golgi retention, and indicates that GOLPH3/3L act on enzymes from a wide-range of Golgi localised modification pathways.

The destabilisation of a wide range of Golgi glycosylation enzymes should perturb the glycosylation status of the cell surface. In order to test this, wild-type and *ΔGOLPH3;ΔGOLPH3L* U2OS cells were probed with a panel of fluorescently-labelled lectins that recognise a range of O- and N-linked glycans (Fig. 4 E). The wild-type and *ΔGOLPH3;ΔGOLPH3L* cells displayed a marked difference in fluorescence intensity for every lectin tested, consistent with the deletion of *GOLPH3* and *GOLPH3L* causing broad-spectrum defects in glycosylation.

GOLPH3 recognises membrane-proximal positively-charged stretches

The identification of a set of high confidence clients for GOLPH3/3L raises the question of what common features they share that allow their recognition. Initially, a biochemical approach was taken to test whether the tails from a range of enzymes were sufficient for binding, thereby excluding the possibility that their retention was indirect by virtue of GOLPH3 binding to an associated protein (McCormick et al., 2000; Hartmann-Fatu et al., 2015; Nilsson et al., 1994). Thus, a series of tail variants of the SI-GFP reporter were generated similar to those used in the Golgi retention assay (Fig. 5 A). The chimeras were overexpressed in HEK293T cells and the cell lysate was subject to affinity chromatography with bacterially-expressed GST-GOLPH3. Immunoblots of the eluate revealed GOLPH3 bound convincingly to chimeras containing tails from several enzymatic pathways including mucin-type O-linked glycosylation (GALNT2, GALNT7, GALNT12), N-linked glycosylation (MGAT2, MANEAL), proteoglycan synthesis (CHSY1, B3GAT3, EXTL3), tyrosine sulfation (TPST2), sialylation (ST6GAL1) and several orphan proteins (GOLM1, CASC4, GOLIM4) (Fig. 5, B and C; and Fig. S4). A few of the tails showed relatively weak binding to GOLPH3 (GALNT4, MGAT1 and MGAT5), or no detectable binding (FUT3), but none of these showed a substantial destabilisation with loss of GOLPH3, or were not detected, and so are less likely to be GOLPH3 clients.

A close examination of the tails that bound GOLPH3 failed to reveal an obvious shared sequence motif. However, all tails shared a short length and the presence of clusters of positively-charged residues (including the amino terminus), with an absence of negatively-charged residues. Conversely, tails which displayed poor or no binding to GOLPH3 displayed a relative paucity of positively-charged clusters and/or the presence of negatively-charged residues. This suggests that GOLPH3 recognises short, positively-charged tails. Moreover, when calculating the predicted net charge of the tails at a cytosolic pH of 7.4, generally only tails with a net charge of ≥ 4+ exhibited robust binding to GOLPH3 *in vitro*. To test this possible charged-based interaction, mutations were made in the tail of SI in an attempt to bestow GOLPH3 binding. Of all the residues targeted, only mutation of a glutamate to alanine was sufficient to induce GOLPH3 binding suggesting that negatively-charged residues do interfere with GOLPH3 recognition (Fig. 5 C). In addition, basic residues were inserted in the membrane proximal region of the tail of SI so that the positive charge was increased without the removal of native SI residues. The insertion of three arginines or three lysines was sufficient to bestow binding; however, the insertion of three histidines, which are not expected to be fully protonated at pH 7.4, was not. Furthermore, the binding increased as the number of positively-charged residues was increased. When the tails were tested in the *in vivo* Golgi retention assay, a

triple arginine or lysine insertion into the tail of SI was sufficient to confer Golgi retention in a GOLPH3+3L-dependent manner, in accordance with the *in vitro* findings, (Fig. 5, D and E). In summary, GOLPH3 appears to be able to recognise its clients by interacting with their short positively-charged tails. This is perhaps best illustrated by the very robust interaction of GOLPH3 with the tail of GALNT2, a tail of only 6 amino acids of which 5 carry a positive-charge (MRRRSR). This simple mode of cargo interaction would explain how GOLPH3 can recognise a broad array of clients *in vivo*.

Bioinformatic analysis of the GOLPH3+3L clientele

The above results suggest that GOLPH3/3L recognise short, positively-charged cytoplasmic tails of type II transmembrane proteins. To see if this correlates with the features of potential GOLPH3/3L clients identified in our two proteomic screens we applied a range of bioinformatic analyses. Since the majority of Golgi-resident clients are known to be type II transmembrane proteins, the data sets were filtered for this topology. Firstly, of the type II proteins that were found in the COPI proteome, we compared those which bound to GOLPH3+3L *in vitro* to those which did not (Fig. 6 A). Logo plot analyses showed that those which bind to GOLPH3/3L have a clear enrichment of basic residues next to the TMD, with blank values dominating further from the TMD, indicating that many of the cytoplasmic tails are not longer than 6-10 residues. Leucine residues are the second most abundant in some positions, consistent with reports that a L-x-x-R/K or L-L-R/K-R/K motifs contribute to binding to GOLPH3 or its yeast ortholog Vps74 (Ali et al., 2012; Tu et al., 2008; Rizzo et al., 2021). We also applied the same analysis to type II proteins that were classified as degraded or non-degraded in the *ΔGOLPH3;ΔGOLPH3L* cell line, and amongst the degraded set there was a very similar enrichment of membrane-proximal positive residues, again followed by blanks indicative of short cytoplasmic tails.

We also compared the prevalence of short, highly basic, cytoplasmic tails in type II membrane proteins from different subcellular locations. Golgi-resident type II proteins clearly have more membrane-proximal positive charges than ER or plasma membrane-residents (Fig. 6 B). Likewise, the type II proteins of the Golgi showed a striking enrichment for shorter cytoplasmic tails compared to those from the ER and plasma membrane (Fig. 6 C). In summary, the proposed GOLPH3+3L-retention signal is greatly over-represented in Golgi type II transmembrane proteins further supporting the case that GOLPH3+3L are major COPI adaptors for intra-Golgi vesicles.

Golgi glycosylation enzymes synthesise glycans in a stepwise manner in which each enzyme adds one or more sugars to generate a structure which is then a substrate for the next enzyme in the pathway. We thus examined the position of the GOLPH3/3L clients in the various pathways in which they act. Each enzyme was categorised as early, intermediate or late-acting based on a recent comprehensive review of glycosylation pathways in human cells (Schjoldager et al., 2020) The glycosylation enzymes that were degraded were almost exclusively early-acting enzymes whereas the non-degraded group contained more intermediate and late-acting enzymes (26% vs 12 %, Fig. 6 D). However, late-acting enzymes were relatively poorly represented across the whole data set

suggesting that many are of low abundance or absent in U2OS cells. Nonetheless, this suggests that glycosylation enzymes that are degraded upon deletion of *GOLPH3* and *GOLPH3L* generally act early in their respective glycosylation pathways.

Discussion

Our global analyses of GOLPH3 and GOLPH3L show that both proteins interact with a wide diversity of Golgi resident enzymes, and for many of these, GOLPH3 and GOLPH3L are required for the retention of the enzyme in the Golgi apparatus. Binding assays with GOLPH3 show that a membrane-proximal cluster of basic residues is sufficient for binding and retention, and this feature, combined with a short cytoplasmic tail, is greatly over-represented in Golgi resident proteins. This suggests that GOLPH3 and GOLPH3L contribute to the Golgi localisation of a wide range of Golgi residents rather than being specific for one particular enzymatic pathway. Nonetheless, it is also clear that other mechanisms can contribute to Golgi enzyme retention, and our data show that retention via direct binding to the COPI coat, or through the TMD, is independent of GOLPH3. This does not preclude some enzymes having multiple retention signals, as is illustrated by ST6GAL1 whose cytoplasmic tail is sufficient for GOLPH3-dependent retention, but which also has a GOLPH3-independent retention signal in its TMD. Such combinations of retention signals could allow precise tuning of the location of an enzyme within the stack, or adjusting of the location between different cell types. It also provides a possible explanation for why the contribution of GOLPH3 to the retention of particular enzymes may have been under-estimated, as removal of GOLPH3 would not cause a complete loss of Golgi retention. Indeed, the early studies on Golgi enzyme retention that identified the role of the TMD also noted a contribution from the cytoplasmic tail which was not pursued at the time (Munro, 1991; Nilsson et al., 1991; Burke et al., 1994, 1992).

How might GOLPH3 and GOLPH3L recognise a wide range of different proteins? Our data, and that of others, highlight the importance of a cluster of basic residues near the start of the TMD. The structure of GOLPH3 shows a conserved acidic patch that covers much of a flat surface on one side of the protein (Fig. 6 E). Thus, one possibility is that when bound into the forming COPI-coated vesicle, GOLPH3 is held close the bilayer with the acidic patch positioned so as to capture short basic tails, and exclude proteins with large, folded, cytoplasmic domains that are likely to be destined for the cell surface. Previous studies on particular GOLPH3 clients have suggested that one or more leucine residues in the tail can also contribute to the interaction (Tu et al., 2008; Ali et al., 2012; Welch and Munro, 2019). This feature is clearly not essential as it is not universal in GOLPH3 clients, but it seems possible that the leucines could either bind at the edge of the acidic patch, or back into the lipid bilayer to optimise the interaction of the basic residues with the acidic patch. A proper understanding of the interaction between the tails and GOLPH3 is likely to require structural analysis similar to the cryo-electron microscopy studies that have revealed how the COPI coat fits onto bilayers (Dodonova et al., 2017; Bykov et al., 2017). It should be noted that quantitative studies have indicated that GOLPH3 is highly abundant in cultured cell lines being present at about 50% of the level of the COPI subunits, suggesting that it could contribute to the recruitment of many Golgi residents into a single COPI-coated vesicle. GOLPH3L is present at only ~10% of the level of GOLPH3 in the cells used in these studies, and mRNA sequencing analysis indicates a similarly lower expression across most tissues (Consortium, 2020; Bekker-Jensen et al., 2017). There have been no reported investigations of its function, apart from a suggestion that it is a negative regulator of

GOLPH3 function (Ng et al., 2013). Our work suggests that it has similar properties to GOLPH3, but our inability to isolate cell lines stably over-expressing the protein does at least suggest some potential negative effect at high levels.

Previous studies on mammalian GOLPH3 have reported roles for the protein that are distinct from Golgi protein retention. In particular, a role in Golgi morphology and exocytosis mediated by an interaction with myosin-18A, and also roles in DNA repair and mTOR signaling (Farber-Katz et al., 2014; Dippold et al., 2009; Rahajeng et al., 2019; Scott et al., 2009). It is of course possible for one protein to have two or more very different functions. However, our GST-GOLPH3 purification did not reveal binding to either myosin-18A, or to the retromer complex that was proposed to be responsible for the effects on mTOR. It is possible that the fusion to GST, or the binding conditions used, interfered with these interactions. However, the BioPlex high-throughput analysis of protein interactions obtained similar findings using C-terminally HA-tagged GOLPH3 and GOLPH3L, both of which gave hits with 4-5 Golgi enzymes, most of which overlapped with our hits, but no hit for either myosin-18A or retromer (Huttlin et al., 2017). In addition, a more recent study of myosin-18A was unable to obtain evidence that it is located on the Golgi (Bruun et al., 2017). Likewise, we were unable to see a difference in the efficiency of cell surface expression of a reporter in the cell line lacking GOLPH3 and GOLPH3L. Consistent with this, a recent study found that immunoglobulin was still efficiently secreted by B-cells from which Myo18A had been removed (Cheung et al., 2021). Further studies will be required to resolve this issue, especially if some of these putative additional roles of GOLPH3 were to prove to be cell-type specific. Nonetheless, we believe that our work, and previous studies in mammalian cultured cells and model organisms provide overwhelming evidence that the major role of GOLPH3 is in Golgi enzyme retention. Interest in the role of GOLPH3 has been increased by the finding that the gene is amplified in a range of solid tumours (Scott et al., 2009; Rizzo et al., 2017; Sechi et al., 2020). A role, or even an exclusive role, for GOLPH3 in retaining enzymes in the Golgi would certainly not be incompatible with these findings as there is extensive evidence that changes in glycosylation are a hallmark of cancer cells, and have been linked to increased tumour growth and invasiveness (Stowell et al., 2015).

The retention of resident proteins in the Golgi apparatus has been investigated for three decades, but progress has been complicated by debate over how secretory cargo proteins move through the Golgi stack. However, recent studies have provided near unequivocal evidence that COPI-coated vesicles selectively recruit Golgi resident proteins rather than cargo (Adolf et al., 2019). Thus, the key question is to understand how the many residents of the Golgi are recruited into these vesicles in different parts of the stack. Our findings demonstrate that GOLPH3 does more than contribute to the retention of a few enzymes, but is rather a major adaptor for cargo sorting. Addressing the precise mechanisms by which it binds COPI and its clients, as well as the contribution of TMDs, direct binding and potentially other adaptors, should now hopefully provide a clear route to answering the long-standing search for understanding of how the Golgi retains its resident proteins as cargo flows past.

Materials and Methods

Plasmids

Please see Table S5 for a full list of the plasmids used in this study. Sequences encoding GOLPH3 (codon optimised to reduce GC content in the N-terminus) and GOLPH3L were synthesised (IDT) and fused at their N-terminus to a TEV protease cleavage site and a GST tag in the vector pOPTG (Olga Perisic) for bacterial expression using the restriction sites NdeI and BamHI.

Plasmids were designed for the transient expression of Golgi enzyme chimeric GFP fusions to serve as baits to test *in vitro* binding to GOLPH3. The short cytoplasmic tail and TMD of sucrase-isomaltase fused to GFP has been used previously as a type II transmembrane plasma membrane reporter (Liu et al., 2018). An N-terminal section of sucrase-isomaltase including the tail, TMD and a short lumenal spacer (comparable to the CTS domains of Golgi enzymes) was PCR-amplified from genomic DNA purified from HeLa cells and fused at its C-terminus to a GAGA linker, a GFP tag and a FLAG tag using the restriction sites NheI, KpnI and NotI in pcDNA3.1+. A host of cytoplasmic tails derived from a variety of Golgi enzymes were used to replace the tail of sucrase-isomaltase in the plasma membrane reporter. The different tails were introduced into the 3' end of forward primers and the chimeras were amplified and cloned into pcDNA3.1+ using the restriction sites NheI and KpnI. Similarly, the TMD of sucrase-isomaltase was replaced with that of ST6GAL1 through DNA fragment synthesis (IDT) and restriction enzyme cloning using the same restriction sites.

For the purpose of the *in vivo* Golgi retention assay, a selection of chimeric fusions where subcloned into a modified bicistronic vector used for the generation of puromycin-resistant cumate-inducible stable cell lines using the restriction sites NheI and NotI. The vector was modified from the PiggyBac vector PBQM812A-1 (System Biosciences) in which the IRES and downstream GFP were removed (John Shin). In order to generate GOLPH3 and GOLPH3L rescue lines; GOLPH3 (kind gift from David Gershlick) and GOLPH3L (PCR amplified from the bacterial expression constructs) were inserted upstream of a chimeric intron, an IRES and mTagBFP2 (synthesised, IDT) by Gibson assembly. Using the restriction site NheI and NotI, these cassettes were inserted into a modified PiggyBac compatible pcDNA3.1+ vector in which the 5' and 3' transposon-specific inverted terminal repeats were inserted upstream of the CMV promoter and downstream of the SV40 poly(A) signal associated with the G418 resistance marker respectively.

Plasmids were designed in order to knockout *GOLPH3* and *GOLPH3L* family genes in mammalian cells using CRISPR-Cas9 gene editing. Oligonucleotides pairs encoding guide RNAs (sgRNAs) targeting specific loci were synthesised with overhangs compatible with the restriction enzyme BbsI and annealed together. BbsI was used to clone the annealed sgRNAs into the bicistronic CRISPR-Cas9 mammalian expression vector pX458 (Feng Zhang) which encodes Cas9-T2A-GFP under a CAG promoter and a U6 promoter driving expression of the sgRNA.

Antibodies

Please see Table S5 for a full list of the antibodies used in this study.

Mammalian cell culture

Human embryonic kidney 293T (ATCC, CRL-3216) and U2OS (ATCC, HTB-96) cells were maintained in a humidified incubator at 37°C with 5% CO_2 in culture medium consisting of Dulbecco's modified Eagle's medium (DMEM, Thermo Fischer Scientific) with 10% fetal bovine serum (FBS, Thermo Fischer Scientific), penicillin-streptomycin and additional selective antibiotics where specified. Furthermore, stable U2OS cell lines expressing Golgi enzyme chimeric reporters under a cumate-inducible promoter were maintained in the presence of 60 µg/ml cumate (System Biosciences). Cells were passaged every 3-4 days in which they were treated with trypsin at 37°C for 2 minutes, resuspended in culture medium and diluted by a factor of 1:10. Cells were regularly screened to confirm they were mycoplasma negative using the MycoAlert kit (Lonza).

Deletion of *GOLPH3* and *GOLPH3L* by CRISPR-Cas9 gene-editing

CRISPR-Cas9 technology was used to simultaneously knockout *GOLPH3* and *GOLPH3L* through the induction of frame-shift mutations and the subsequent introduction of premature stop codons in early constitutive exons. *GOLPH3* was targeted at one site in exon 2 (target sequence, 5'-GAGAGGAAGGTTACAACTAG-3') in order to induce small indel mutations while two sites 63 base pairs apart were targeted to introduce a larger out-of-frame deletion mutation in exon 2 of *GOLPH3L* (target sequence 1, 5'-CTTCTTCCATAAGAGTAAGG-'3; target sequence 2, 5'-GTAATGCAGTTAGGTTTGCT-'3). Wild-type U2OS were seeded at a density of $2x10^4$ cells/cm^2 in T-75 flasks in culture medium in a humidified incubator at 37°C with 5% CO_2. Once cells were between 50-80% confluent, cells were transfected with a total of 15 µg of DNA of a bicistronic plasmid encoding the sgRNAs and Cas9-T2A-EGFP. 1 mg/mL polyethylenimine (PEI, Polyscience) in PBS was used for the transfection at a ratio of 3:1 (µL:µg) with DNA in which PEI was incubated in Opti-MEM (Thermo Fischer Scientific) for 5 minutes prior to mixing with the DNA. DNA complexes were subsequently incubated for a further 15 minutes prior to dropwise addition to cells. 24 hours after transfection, GFP-positive cells were sorted to one cell per well into 96-well plates containing fresh culture medium using a MoFlo Cell Sorter (Beckman Coulter) and clones were gradually expanded to 6-well format over the course of several weeks. Whole cell lysates of clones were analysed by western blot to confirm the absence of the protein of interest and candidate knockout clones were validated by genotyping PCR. Furthermore, the proteome of the final *ΔGOLPH3;ΔGOLPH3L* U2OS candidate clone was analysed by mass spectrometry to confirm the cell line was a true knockout (described below).

PiggyBac transposon stable cell line generation

The PiggyBac Transposon system (System Biosciences) was used to generate stable cell lines expressing either GFP-tagged Golgi enzyme chimeric reporters under a cumate-inducible promoter or *GOLPH3L* and *GOLPH3* under a CMV promoter. Wild-type and *ΔGOLPH3;ΔGOLPH3L* U2OS cells

were seeded in 6-well plates at a density of $2x10^4$ cells/cm^2 in culture medium in a humidified incubator at 37°C with 5% CO_2. After cells reached 50% confluency, they were transfected with 0.5 µg of a PiggyBac-compatible expression vector and 0.2 µg of PiggyBac transposase (PB210PA-1). 48 hours after transfection, cells were expanded to T-75 flasks and 72 hours after transfection, cells were subject to selection in culture medium with 0.5 µg/mL puromycin (cumate-inducible GFP-tagged reporter cell lines) or 200 µg/mL G418 (GOLPH3 and GOLPH3L rescue cell lines). Cells were cultured under selection for approximately 2 or 3 passages to ensure robust selection. Where stated, GOLPH3 and GOLPH3L rescue cell lines were also subject to cloning by limiting dilution into 96-well plates. Selection was maintained throughout expansion and the resulting clones were validated by western blot to select lines with moderate to high expression of the gene of interest where possible.

GST pulldowns

BL21-CodonPlus(DE3)-RIL competent cells (Agilent) were transformed with constructs encoding GST-GOLPH3, GST-GOLPH3L or a GST alone and cells were plated on 2xTY agar plates containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and left overnight at 37°C. Single colonies were selected for inoculation of 25 mL overnight liquid cultures of 2xTY containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol at 37°C at 220 rpm. Starter cultures were used to inoculate larger cultures at a ratio of 1:20 and they were incubated until they reached an OD_{600} of 0.5 - 0.8. Cultures were induced with 100 µM Isopropyl β-D-1-thiogalactopyranoside overnight at 16°C. Cells were pelleted by centrifugation at 4500 x g for 15 minutes at 4°C and washed once with ice cold PBS by resuspension and centrifugation. Bacterial cells were resuspended in lysis buffer consisting of 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 1 mM PMSF (Sigma) and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). Cells were sonicated on ice for 1 minute; 10 seconds on, 10 seconds off at 45% amplitude using a Sonic Vibra-Cell lance sonicator. Cells were placed on fresh ice for at least 5 minutes and incubated with agitation at 4°C for a further 10 minutes. Lysates were clarified by centrifugation at 32,000 x *g* for 10 minutes at 4°C. Glutathione sepharose 4B beads (GE Life Sciences) were washed with lysis buffer by resuspension and pelleting by centrifugation at 100 x *g* for 1 minute. Clarified bacterial lysates were mixed with the glutathione beads and incubated with agitation at 4°C for 30 minutes. Beads were subjected to one wash with lysis buffer, one high salt wash (lysis buffer with 500 mM NaCl) and another four lysis buffer washes. Loaded beads were kept on ice prior to addition of prey lysates.

Where GST pulldown samples were destined for downstream mass spectrometry analysis, 4x T-175 flasks of wild-type HEK293T cells per bait were grown to confluency. Cells were harvested by resuspension in culture medium and residual cells were recovered from flasks using an EDTA solution wash. Cell suspensions were pelleted by centrifugation at 300 x *g* for 5 minutes at 4°C and washed once in ice cold PBS by resuspension and centrifugation. Cells were resuspended in lysis buffer and sonicated for only 10 seconds at 45% amplitude using a Sonic Vibra-Cell lance sonicator. Lysates were clarified by centrifugation at 16,100 x *g* for 10 minutes at 4°C and the supernatant was mixed with the preloaded beads and incubated with agitation for 1-2 hours at 4°C. Beads were

washed 5 times with lysis buffer and specific interactors were eluted in lysis buffer with 1.5 M NaCl. High salt elutions were subject to TCA/acetone reprecipitation and were resolubilised in 1x LDS with 10% β-mercaptoethanol (BME). Bait proteins were eluted by boiling in 2x LDS with 10% BME.

For GST pulldown experiments involving GFP-tagged chimeric reporters, HEK293T cells were seeded in T-75 flasks in culture medium. Once cells reached 50-80% confluency, cells in each flask were transfected with 15 µg of plasmid DNA encoding the chimeric reporters using PEI as described previously. 48-72 hours after transfection, cells were harvested, washed, lysed as for samples destined for use in downstream mass spectrometry analysis. GST-GOLPH3 loaded beads were split evenly amongst the different reactions and mixed with the lysates containing the different chimeric GFP-tagged reporters and incubated with agitation for 1 hour at 4°C. A fraction of the clarified lysate was retained for the purpose of input controls. Beads were washed 5 times with lysis buffer and proteins were eluted by boiling in 2x LDS with 10% BME or 50 mM TCEP pH 7.0.

Mammalian cell lysis

Wild-type, *ΔGOLPH3;ΔGOLPH3L* and rescue U2OS cell lines were seeded at a density of $2x10^4$ cells/cm^2 in 6-well plates or 100 mm dishes in the selection medium or culture medium. Once cells reached 80-90% confluency, cells required for western blotting were washed once with EDTA solution, incubated in trypsin solution for 2 minutes at 37°C and resuspended in culture medium. Cells required for whole cell proteomic analysis by mass spectrometry were washed once with ice cold PBS and detached from flasks by scraping in PBS. All cell suspensions were pelleted by centrifugation at 300 x *g* for 5 minutes at 4°C and washed once in ice cold PBS by resuspension and centrifugation. Cells for western blotting were resuspended in lysis buffer while cells required for mass spectrometry were lysed in 8M urea, 20 mM Tris HCl. All cells were sonicated for 1 minute using a Misonix 300 water sonicator for 1 minute 10 seconds on, 10 seconds off at amplitude 5.0. Lysates were cleared by centrifugation at 16,100 x *g* for 10 minutes at 4°C. The protein concentration of the lysates was measured using a Pierce BCA Protein Assay Kit (Thermo Fischer Scientific) and Infinite F200 plate reader (Tecan). The protein samples for western blotting were normalised across treatments and were mixed with loading dye and reducing agent to a final concentration of 1x LDS, 50 mM TCEP pH 7.0. Protein samples for mass spectrometry were diluted to 2 µg/mL and snap frozen in liquid nitrogen and stored at -80°C until required.

SDS-PAGE and immunoblotting

Protein samples were incubated at 90°C for 3 minutes and loaded into Novex 4-20% Tris-Glycine Mini Gels (Thermo Fischer Scientific) and resolved for 1 hour at a constant voltage of 175 V in tris-glycine SDS running buffer. Total protein was stained for in which gels were incubated in InstantBlue Coomassie stain (Expedeon) for 1 hour to overnight at room temperature with agitation. Gels were washed five times for 5 minutes in H_2O prior to imaging. Alternatively, gels were subjected to a western blot in which protein was transferred onto a 0.45 µm nitrocellulose membrane using a Mini Trans-Blot Cell (Bio-Rad), in transfer buffer in the presence of an ice block for 1 hour at a constant

current of 255 mA. Blots were blocked in 3% w/v non-fat dry milk in PBST (0.1% Tween-20 in PBS) for 1 hour at room temperature with agitation. Blots were incubated with the primary antibody diluted in 3% milk in PBST overnight at 4°C with agitation. Blots were washed four times for 5 minutes in PBST and incubated with the secondary antibody diluted in 3% milk in PBS for 1 hour at room temperature with agitation. Blots were washed four times for 5 minutes in PBST. Where applicable, chemiluminescent substrates (Amersham ECL or Amersham ECL Prime, Cytiva) were added to blots 3 minutes prior to exposure to X-ray films which were developed using a JP-33 film processer (JPI Healthcare Solutions). Alternatively, blots were imaged using a ChemiDoc (Bio-Rad). Where specified, blots stained with an AF555-conjugated secondary antibody were also visualised using a ChemiDoc.

Lectin binding

Wild-type and *ΔGOLPH3;ΔGOLPH3L* U2OS cell lines were seeded at a density of $2x10^4$ cells/cm^2 in T-75 flasks in culture medium in a humidified incubator at 37°C with 5% CO_2. Once cells reached 80-90% confluency, they were washed once with EDTA solution and incubated in accutase (Sigma) for 2 minutes at 37°C. Cells were resuspended in ice cold FACS buffer (2% FBS in PBS) and transferred into a round-bottomed 96-well plate at approximately 10^6 cells/well. Cells were washed once in which cell suspensions were centrifuged at 300 x g for 5 minutes, the supernatant removed and cells resuspended in FACS buffer. Cells were incubated with a panel of 7 fluorescein-labelled lectins (final concentration 20 µg/mL, Vector Biolabs) and a fixable viability dye eFluor 780 (1:1000, Thermo Fischer Scientific) diluted in FACS buffer on ice in darkness for 30 minutes. As and where specified, controls to validate lectin specificities were included in which lectins were preincubated in FACS buffer containing saturating concentrations of competitive sugars at least 30 minutes prior to addition to cells. Cells were washed 3 times in FACS buffer and fixed in 4% paraformaldehyde (PFA) in PBS for 20 minutes at room temperature. Cells were washed a further two times in FACS buffer and kept at 4°C in darkness until required. Cell suspensions were filtered using a 100 µm plate filter immediately prior to analysis on an LSRII flow cytometer (BD Biosciences) or an EC800 flow cytometer (Sony). Data was analysed and histogram plots generated using FlowJo V10. Singlets were gated according to forward and side scatter profiles, dead cells were excluded from analysis using the viability stain. Single colour control samples were included to confirm the appropriate compensation parameters.

Golgi retention assay

Inducible stable cell lines expressing GFP-tagged Golgi enzyme chimeric reporters were cultured in 6-well plate format in selection media containing 60 µg/mL cumate for at least a week prior to analysis. Once cells reached 80-90% confluency, they were washed once with EDTA solution and incubated in accutase for 2 minutes at 37°C. Cells were resuspended in selection media and transferred into a deep 96-well plate. Cells were washed once by centrifugation at 300 x g for 5 minutes followed by resuspension in ice cold FACS buffer. Cells were transferred to a round-bottomed 96-well plate and were incubated with an AF647-conjugated anti-GFP antibody (1:20, BioLegend) and fixable viability

dye eFluor 780 (1:1000, BD Biosciences) diluted in FACS buffer on ice in darkness for 30 minutes. Cells were subsequently washed, fixed and analysed as described for lectin stains. Furthermore, GFP negative cells were excluded from analysis and a ratio of the AF647 signal to the GFP signal was used to derive a quantitative parameter for Golgi retention.

Immunofluorescence

Cells were seeded onto multispot microscope slides (Hendley-Essex) in culture medium in a humidified incubator at 37°C with 5% CO_2. 24 hours after seeding, cells were washed twice in PBS and fixed in 4% PFA in PBS for 20 minutes at room temperature. Cells were washed twice in PBS and permeabilised for 10 minutes in 10% Triton X-100. Cells were washed 5 times in PBS and blocked for 1 hour in blocking buffer (20% FBS, 1 % Tween-20 in PBS). Blocking buffer was aspirated and cells were incubated in the primary antibody cocktail diluted in blocking buffer for 1 hour. Cells were washed twice in PBS, incubated in blocking buffer for 10 minutes and washed twice again in PBS prior to incubation in the secondary antibody cocktail diluted in blocking buffer for 1 hour. Cells were washed twice in PBS, incubated in blocking buffer for 10 minutes and washed twice again. PBS was aspirated and VECTASHIELD mounting media (Vector Biolabs) was added to cells prior to the addition of the coverslip. The coverslip was sealed with nail varnish and slides were imaged using a Leica TCS SP8 confocal microscope. Images were processed using Adobe Photoshop CC 2017.

Mass spectrometry

Protein digestion

Protein samples (10 x 200 ug each) in lysis buffer (8M urea, 20 mM Tris pH8) were reduced with 5 mM DTT at 56°C for 30 min and alkylated with 10 mM iodoacetamide in the dark at room temperature for 30 min. The samples were then diluted to 4M urea and digested with Lys-C (Promega), 67:1 (protein: Lys-C ratio, w/w) for 4 hr at 25°C. Next, the samples were further diluted to 1.6 M urea and were digested with trypsin (Promega) 50:1 (protein: trypsin ratio, w/w) over night, at 25°C. Digestion was stopped by the addition of formic acid (FA) to a final concentration of 0.5%. Any precipitates were removed by centrifugation at 13000 rpm for 8 min. The supernatants were desalted using a home-made C18 stage tips (3M Empore) contained 4 mg of Poros R3 (Applied Biosystems) resin. Bound peptides were eluted with 30-80% acetonitrile (MeCN) in 0.1% TFA and lyophilized.

Tandem mass tag (TMT) labeling

Peptide mixtures from each condition was re-suspended in 74 ul of 200mM Hepes, pH8.3. TMT 10plex reagent (Thermo Fisher Scientific), 36 ul (720 ug) was added and incubated at room temperature for an hour. The labeling reaction was then terminated by incubation with 7.3 ul 5% hydroxylamine. The labeled peptides were pooled into a single sample and was desalted using the same stage tips method as above.

Off-line High pH reverse-phase peptides fractionation

About 200ug of the labeled peptides were separated on an off-line, high pressure liquid

chromatography (HPLC). The experiment was carried out using XBridge BEH130 C18, 5 µm, 2.1 x 150mm (Waters) column with XBridge BEH C18 5 µm Van Guard cartridge, connected to an Ultimate 3000 Nano/Capillary LC System (Dionex). Peptides were separated with a gradient of 1-90% B (A: 5% MeCN/10 mM ammonium bicarbonate, pH8; B: MeCN/10mM ammonium bicarbonate, pH8, [9:1]) in 60 min at a flow rate of 250 µl/min. A total of 54 fractions were collected, they were combined into 18 fractions and lyophilized. Dried peptides were resuspended in 1% MeCN/0.5% FA and desalted using stage tips for mass spectrometry analysis.

Mass spectrometry analysis

Peptides were separated on an Ultimate 3000 RSLC nano System (Thermo Scientific), using a binary gradient consisting of buffer A (2% MeCN, 0.1% formic acid) and buffer B (80% MeCN, 0.1% formic acid). Eluted peptides were introduced directly via a nanospray ion source into a Q Exactive Plus hybrid quardrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in standard data dependent mode, performed survey full-scan (MS, m/z = 380-1600) with a resolution of 70000, followed by MS2 acquisitions of the 15 most intense ions with a resolution of 35000 and NCE of 33%. MS target values of 3e6 and MS2 target values of 1e5 were used. Dynamic exclusion was enabled for 40s.

GST affinity chromatography mass spectrometry

Gel samples were destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM DTT, and alkylated with 55 mM iodoacetamide. Digestion was with 6 ng/µl trypsin (Promega, UK) overnight at 37°C, and peptides extracted in 2% v/v formic acid 2% v/v acetonitrile, and analysed by nano-scale capillary LC-MS/MS (Ultimate U3000 HPLC, Thermo Scientific Dionex) at a flow of ~ 300 nL/min. A C18 Acclaim PepMap100 5 µm, 100 µm x 20 mm nanoViper (Thermo Scientific Dionex), trapped the peptides prior to separation on PicoChip Column: 75um ID x 15um tip packed with 105mm 3um Reprosil-PUR C18-AQ 120A (New Objective). Peptides were eluted with an acetonitrile gradient. The analytical column outlet was interfaced via a nano-flow electrospray ionisation source with a linear ion trap mass spectrometer (Orbitrap Velos, Thermo Scientific). Data dependent analysis was performed using a resolution of 30,000 for the full MS spectrum, followed by ten MS/MS spectra in the linear ion trap. MS spectra were collected over a m/z range of 300–2000. MS/MS scans were collected using a threshold energy of 35 for collision-induced dissociation. LC-MS/MS data were searched against the UniProt KB database using Mascot (Matrix Science), with a precursor tolerance of 10 ppm and a fragment ion mass tolerance of 0.8 Da. Two missed enzyme cleavages and variable modifications for oxidised methionine, carbamidomethyl cysteine, pyroglutamic acid, phosphorylated serine, threonine and tyrosine were included. MS/MS data were validated using the Scaffold programme (Proteome Software Inc).

Mass spectrometry data analysis

The acquired MSMS raw files were processed using MaxQuant (Cox and Mann, 2008), with the integrated Andromeda search engine (v.1.6.6.0). MSMS spectra were searched against *Homo*

sapiens UniProt Fasta database. Carbamidomethylation of cysteines was set as fixed modification, while methionine oxidation and N-terminal acetylation (protein) were set as variable modifications. Protein quantification requires 1 (unique+ razor) peptide. Other parameters in MaxQuant were set to default values. MaxQuant output file, proteinGroups.txt was then processed with Perseus software (v 1.6.6.0). After uploading the matrix, the data was filtered, to remove identifications from reverse database, modified peptide only, and common contaminants.

Bioinformatics

Type II transmembrane (TM) proteins and their TM span locations were initially identified from the reviewed, non-redundant UniProt entries for the human proteome (The UniProt Consortium, 2021). Entries were manually reviewed to correct obvious errors, which mostly related to sub-cellular localisation and signal peptide annotation. TM span edges were then refined using a single, consistent approach employed previously (Parsons et al., 2019). In précis, this considered positions ±5 residues from the stated UniProt TM edge, found the point of maximum hydrophobicity difference between the five preceding and five subsequent residues, and then trimmed the end residue if hydrophilic (here using Arg, Lys, Asp, Glu, Gln, Asn, His, or Ser) or extended it if a hydrophobic residue is next (Phe, Met, Ile, Leu, Val, Cys, Trp, Ala, Thr, or Gly). Subsequently, the TM protein entries were matched to the protein IDs used in various analysis groups via the gene name to accession code mapping ("gene2acc") at UniProt, which ties redundant protein entries to their gene of origin.

Logo plots were generated using the Python script available at github.com/tjs23/logo_plot/. Inputs to the plots were one-letter protein sequences of TM spans, aligned on their first residue, with flanking regions. These regions covered positions from 15 residues before to 35 residues after each TM start position (the N-terminal edge), which also acted as the anchor point to compare the TM spans from different proteins. Where TM spans had short flanking tails that did not reach the edge of the plot regions, the ends of the protein sequences were padded with "X", which was plotted with the real amino acid types. This proved helpful to illustrate the occurrence of short tails.

From the curated set of Type II TM proteins, those with a known, unambiguous sub-cellular localisation within the membranes of the endoplasmic reticulum, Golgi apparatus or plasma membrane were selected for analysis of positively charged near-TM groups. These groups included arginine side chains, lysine side chains and/or an N-terminal α-amino group. Counts were made for the occurrence of these within a six residue region just outside the TM span; from the cytoplasmic TM edge, as described above.

Supplemental material

Figure S1. Validation of the *ΔGOLPH3;ΔGOLPH3L* U2OS cell line.

Figure S2. Flow cytometry gating strategies.

Figure S3. Genetic rescue of the destablisation of Golgi residents in *ΔGOLPH3;ΔGOLPH3L* cells.

Figure S4. GOLPH3 interacts with the tails of GALNT2 and GALNT7.

Table S1. GST affinity chromatography data for GOLPH3 and GOLPH3L.

Table S2. Comparison of affinity chromatography data to COPI vesicle proteome.

Table S3. Proteomic data comparing wild-type and *ΔGOLPH3;ΔGOLPH3L* cells.

Table S4. Lists of proteins used for bioinformatic analyses.

Table S5. Antibodies and plasmids.

Acknowledgements

We thank David Owen, Jonathan Kaufman and Jasmine Cornish for discussions on the structure of GOLPH3. Thanks to Ryan Britnell for help with analysis of the proteomics data. Thanks to Maria Daly and Fan Zhang for help with cell sorting and flow cytometry advice. GOLPH3 cDNA was a kind gift from David Gershlick. Funding was from the Medical Research Council, as part of United Kingdom Research and Innovation (also known as UK Research and Innovation), (file reference number MC_U105178783).

Author contributions

Conceptualisation: L.G. Welch and S. Munro. Formal analysis: L.G. Welch. Funding acquisition and supervision: S. Munro. Investigation and methodology: L.G. Welch, S-Y. Peak-Chew F. Begum. Visualisation: L.G. Welch and T.J. Stevens. Writing – original draft: L.G. Welch. Writing – review and editing: S. Munro.

References

Adolf, F., M. Rhiel, B. Hessling, Q. Gao, A. Hellwig, J. Bethune, and F.T. Wieland. 2019. Proteomic profiling of mammalian COPII and COPI vesicles. *Cell Rep*. 26:250–265. doi:10.1016/j.celrep.2018.12.041.

Ali, M.F., V.B. Chachadi, A. Petrosyan, and P.W. Cheng. 2012. Golgi phosphoprotein 3 determines cell binding properties under dynamic flow by controlling Golgi localization of core 2 N-acetylglucosaminyltransferase 1. *J Biol Chem*. 287:39564–39577. doi:10.1074/jbc.M112.346528.

Bekker-Jensen, D.B., C.D. Kelstrup, T.S. Batth, S.C. Larsen, C. Haldrup, J.B. Bramsen, K.D. Sørensen, S. Høyer, T.F. Ørntoft, C.L. Andersen, M.L. Nielsen, and J. v. Olsen. 2017. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems*. 4:587-599.e4. doi:10.1016/j.cels.2017.05.009.

Bretscher, M.S., and S. Munro. 1993. Cholesterol and the Golgi apparatus. *Science*. 261:1280–1281. doi:10.1126/science.8362242.

Bruun, K., J.R. Beach, S.M. Heissler, K. Remmert, J.R. Sellers, and J.A. Hammer. 2017. Re-evaluating the roles of myosin 18Aα and F-actin in determining Golgi morphology. *Cytoskeleton*. 74:205–218.

Burke, J., J.M. Pettitt, D. Humphris, and P.A. Gleeson. 1994. Medial-Golgi retention of N-acetylglucosaminyltransferase-I - contribution from all domains of the enzyme. *J Biol Chem*. 269:12049–12059.

Burke, J., J.M. Pettitt, H. Schachter, M. Sarkar, and P.A. Gleeson. 1992. The transmembrane and flanking sequences of β1,2-N-acetylglucosaminyltransferase I specify medial-Golgi localization. *J Biol Chem*. 267:24433–24440.

Bykov, Y.S., M. Schaffer, S.O. Dodonova, S. Albert, J.M. Plitzko, W. Baumeister, B.D. Engel, and J.A. Briggs. 2017. The structure of the COPI coat determined within the cell. *Elife*. 6:e32493. doi:10.7554/eLife.32493.

Chang, W.L., C.W. Chang, Y.Y. Chang, H.H. Sung, M.D. Lin, S.C. Chang, C.H. Chen, C.W. Huang, K.S. Tung, and T.B. Chou. 2013. The *Drosophila* GOLPH3 homolog regulates the biosynthesis of heparan sulfate proteoglycans by modulating the retrograde trafficking of exostosins. *Development*. 140:2798–2807. doi:10.1242/dev.087171.

Cheung, M.B., G. Enyindah-Asonye, K. Matsui, I. Kosik, N. Dvorina, W.M. Baldwin, J.W. Yewdell, and N. Gupta. 2021. Cutting Edge: Myosin 18A Is a Novel Checkpoint Regulator in B Cell Differentiation and Antibody-Mediated Immunity. *The Journal of Immunology*. 206:2521–2526.

Consortium, Gte. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, NY)*. 369:1318–1330.

Cox, J., and M. Mann. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*. 26:1367–1372. doi:10.1038/nbt.1511.

Dippold, H.C., M.M. Ng, S.E. Farber-Katz, S.K. Lee, M.L. Kerr, M.C. Peterman, R. Sim, P.A. Wiharto, K.A. Galbraith, S. Madhavarapu, G.J. Fuchs, T. Meerloo, M.G. Farquhar, H. Zhou, and S.J.

Field. 2009. GOLPH3 bridges phosphatidylinositol-4-phosphate and actomyosin to stretch and shape the Golgi to promote budding. *Cell*. 139:337–351. doi:10.1016/j.cell.2009.07.052.

Dodonova, S.O., P. Aderhold, J. Kopp, I. Ganeva, S. Rohling, W.J.H. Hagen, I. Sinning, F. Wieland, and J.A.G. Briggs. 2017. 9Å structure of the COPI coat reveals that the Arf1 GTPase occupies two contrasting molecular environments. *Elife*. 6:e26691. doi:10.7554/eLife.26691.

Dunlop, M.H., A.M. Ernst, L.K. Schroeder, D.K. Toomre, G. Lavieu, and J.E. Rothman. 2017. Land-locked mammalian Golgi reveals cargo transport between stable cisternae. *Nature communications*. 8:432.

Eckert, E.S., I. Reckmann, A. Hellwig, S. Rohling, A. El-Battari, F.T. Wieland, and V. Popoff. 2014. Golgi phosphoprotein 3 triggers signal-mediated incorporation of glycosyltransferases into coatomer-coated (COPI) vesicles. *J Biol Chem*. 289:31319–31329. doi:10.1074/jbc.M114.608182.

Farber-Katz, S.E., H.C. Dippold, M.D. Buschman, M.C. Peterman, M. Xing, C.J. Noakes, J. Tat, M.M. Ng, J. Rahajeng, D.M. Cowan, G.J. Fuchs, H. Zhou, and S.J. Field. 2014. DNA Damage Triggers Golgi Dispersal via DNA-PK and GOLPH3. *Cell*. 156. doi:10.1016/j.cell.2013.12.023.

Flynn, R.A., K. Pedram, S.A. Malaker, P.J. Batista, B.A.H. Smith, A.G. Johnson, B.M. George, K. Majzoub, P.W. Villalta, J.E. Carette, and C.R. Bertozzi. 2021. Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell*. 184:1–16. doi:10.1016/j.cell.2021.04.023.

van Galen, J., F. Campelo, E. Martinez-Alonso, M. Scarpa, J.A. Martinez-Menarguez, and V. Malhotra. 2014. Sphingomyelin homeostasis is required to form functional enzymatic domains at the trans-Golgi network. *J Cell Biol*. 206:609–618. doi:10.1083/jcb.201405009.

Glick, B.S., and A. Nakano. 2009. Membrane traffic within the Golgi apparatus. *Annual review of cell and developmental biology*. 25:113–132.

Gomez-Navarro, N., and E. Miller. 2016. Protein sorting at the ER-Golgi interface. *Journal of Cell Biology*. 216:769–778. doi:10.1083/jcb.201610031.

Hartmann-Fatu, C., F. Trusch, C.N. Moll, I. Michin, A. Hassinen, S. Kellokumpu, and P. Bayer. 2015. Heterodimers of tyrosylprotein sulfotransferases suggest existence of a higher organization level of transferases in the membrane of the trans-Golgi apparatus. *Journal of Molecular Biology*. 427:1404–1412. doi:10.1016/j.jmb.2015.01.021.

Huttlin, E.L., R.J. Bruckner, J.A. Paulo, J.R. Cannon, L. Ting, K. Baltier, G. Colby, F. Gebreab, M.P. Gygi, H. Parzen, J. Szpyt, S. Tam, G. Zarraga, L. Pontano-Vaites, S. Swarup, A.E. White, D.K. Schweppe, R. Rad, B.K. Erickson, R.A. Obar, K.G. Guruharsha, K. Li, S. Artavanis-Tsakonas, S.P. Gygi, and J.W. Harper. 2017. Architecture of the human interactome defines protein communities and disease networks. *Nature*. 545. doi:10.1038/nature22366.

Isaji, T., S. Im, W. Gu, Y. Wang, Q. Hang, J. Lu, T. Fukuda, N. Hashii, D. Takakura, N. Kawasaki, H. Miyoshi, and J. Gu. 2014. An oncogenic protein Golgi phosphoprotein 3 up-regulates cell migration via sialylation. *J Biol Chem*. 289:20694–20705. doi:10.1074/jbc.M113.542688.

Jayaprakash, N.G., and A. Surolia. 2017. Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal*. 474. doi:10.1042/BCJ20170111.

Kingsley, D.M., K.F. Kozarsky, M. Segal, and M. Krieger. 1986. Three types of low density lipoprotein receptor-deficient mutant have pleiotropic defects in the synthesis of N-linked, O-linked, and lipid-linked carbohydrate chains. *The Journal of Cell Biology*. 102. doi:10.1083/jcb.102.5.1576.

Kornfeld, R., and S. Kornfeld. 1985. Assembly of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry*. 54:631–664. doi:10.1146/annurev.bi.54.070185.003215.

Liu, L., B. Doray, and S. Kornfeld. 2018. Recycling of Golgi glycosyltransferases requires direct binding to coatomer. *Proc Natl Acad Sci U S A*. 115:8984–8989. doi:10.1073/pnas.1810291115.

Lombard, V., H. Golaconda Ramulu, E. Drula, P.M. Coutinho, and B. Henrissat. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42:D490–D495. doi:10.1093/nar/gkt1178.

Lujan, P., and F. Campelo. 2021. Should I stay or should I go? Golgi membrane spatial organization for protein sorting and retention. *Archives of Biochemistry and Biophysics*. 707. doi:10.1016/j.abb.2021.108921.

Maccioni, H.J.F., R. Quiroga, and M.L. Ferrari. 2011. Cellular and molecular biology of glycosphingolipid glycosylation. *Journal of neurochemistry*. 117:589–602.

McCormick, C., G. Duncan, K.T. Goutsos, and F. Tufaro. 2000. The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proc Natl Acad Sci U S A*. 97:668–673. doi:10.1073/pnas.97.2.668.

Moremen, K.W., M. Tiemeyer, and A. V Nairn. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 13:448–462. doi:10.1038/nrm3383.

Munro, S. 1991. Sequences within and adjacent to the transmembrane segment of α-2,6-sialyltransferase specify Golgi retention. *EMBO J*. 10:3577–3588. doi:DOI 10.1002/j.1460-2075.1991.tb04924.x.

Ng, M.M., H.C. Dippold, M.D. Buschman, C.J. Noakes, and S.J. Field. 2013. GOLPH3L antagonizes GOLPH3 to determine Golgi morphology. *Molecular Biology of the Cell*. 24:796–808. doi:10.1091/mbc.E12-07-0525.

Nilsson, T., M.H. Hoe, P. Slusarewicz, C. Rabouille, R. Watson, F. Hunte, G. Watzele, E.G. Berger, and G. Warren. 1994. Kin recognition between medial Golgi enzymes in HeLa cells. *EMBO Journal*. 13:562–574. doi:10.1002/j.1460-2075.1994.tb06294.x.

Nilsson, T., J.M. Lucocq, D. Mackay, and G. Warren. 1991. The membrane spanning domain of β-1,4-galactosyltransferase specifies trans Golgi localization. *EMBO J*. 10:3567–3575. doi:10.1002/j.1460-2075.1991.tb04923.x.

Pantazopoulou, A., and B.S. Glick. 2019. A kinetic view of membrane traffic pathways can transcend the classical view of Golgi compartments. *Frontiers in Cell and Developmental Biology*. 7:153. doi:10.3389/fcell.2019.00153.

Parsons, H.T., T.J. Stevens, H.E. McFarlane, S. Vidal-Melgosa, J. Griss, N. Lawrence, R. Butler, M.M.L. Sousa, M. Salemi, W.G.T. Willats, C.J. Petzold, J.L. Heazlewood, and K.S. Lilley. 2019. Separating Golgi Proteins from *Cis* to *Trans* Reveals Underlying Properties of Cisternal Localization. *The Plant Cell*. 31. doi:10.1105/tpc.19.00081.

Pascoal, C., R. Francisco, T. Ferro, V. dos Reis Ferreira, J. Jaeken, and P.A. Videira. 2020. CDG and immune response: From bedside to bench and back. *Journal of Inherited Metabolic Disease*. 43:90–124. doi:10.1002/jimd.12126.

Pereira, N.A., H.X. Pu, H. Goh, and Z. Song. 2014. Golgi phosphoprotein 3 mediates the Golgi localization and function of protein O-linked mannose β-1,2-N-acetlyglucosaminyltransferase 1. *J Biol Chem*. 289:14762–14770. doi:10.1074/jbc.M114.548305.

Pinho, S.S., and C.A. Reis. 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 15:540–555. doi:10.1038/nrc3982.

Rahajeng, J., R.S. Kuna, S.L. Makowski, T.T.T. Tran, M.D. Buschman, S. Li, N. Cheng, M.M. Ng, and S.J. Field. 2019. Efficient Golgi Forward Trafficking Requires GOLPH3-Driven, PI4P-Dependent Membrane Curvature. *Developmental Cell*. 50:573-585.e5. doi:10.1016/j.devcel.2019.05.038.

Rizzo, R., S. Parashuraman, G. D'Angelo, and A. Luini. 2017. GOLPH3 and oncogenesis: What is the molecular link? *Tissue Cell*. 49:170–174. doi:10.1016/j.tice.2016.06.008.

Rizzo, R., D. Russo, K. Kurokawa, P. Sahu, B. Lombardi, D. Supino, M.A. Zhukovsky, A. Vocat, P. Pothukuchi, V. Kunnathully, L. Capolupo, G. Boncompain, C. Vitagliano, F. Zito Marino, G. Aquino, D. Montariello, P. Henklein, L. Mandrich, G. Botti, H. Clausen, U. Mandel, T. Yamaji, K. Hanada, A. Budillon, F. Perez, S. Parashuraman, Y.A. Hannun, A. Nakano, D. Corda, G. D'Angelo, and A. Luini. 2021. Golgi maturation-dependent glycoenzyme recycling controls glycosphingolipid biosynthesis and cell growth via GOLPH3. *The EMBO Journal*. e107238. doi:10.15252/embj.2020107238.

Sandhoff, R., and K. Sandhoff. 2018. Emerging concepts of ganglioside metabolism. *FEBS letters*. 592:3835–3864.

Schjoldager, K.T., Y. Narimatsu, H.J. Joshi, and H. Clausen. 2020. Global view of human protein glycosylation pathways and functions. *Nature Reviews Molecular Cell Biology*. 21:729–749. doi:10.1038/s41580-020-00294-x.

Schmitz, K.R., J. Liu, S. Li, T.G. Setty, C.S. Wood, C.G. Burd, and K.M. Ferguson. 2008. Golgi localization of glycosyltransferases requires a Vps74p oligomer. *Dev Cell*. 14:523–534. doi:10.1016/j.devcel.2008.02.016.

Scott, H., and V.M. Panin. 2014. N-Glycosylation in Regulation of the Nervous System.

Scott, K.L., O. Kabbarah, M.C. Liang, E. Ivanova, V. Anagnostou, J. Wu, S. Dhakal, M. Wu, S. Chen, T. Feinberg, J. Huang, A. Saci, H.R. Widlund, D.E. Fisher, Y. Xiao, D.L. Rimm, A. Protopopov, K.K. Wong, and L. Chin. 2009. GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. *Nature*. 459:1085–1090. doi:10.1038/nature08109.

Sechi, S., A. Frappaolo, G. Belloni, G. Colotti, and M.G. Giansanti. 2015. The multiple cellular functions of the oncoprotein Golgi phosphoprotein 3. 6. 3493–3506 pp.

Sechi, S., A. Frappaolo, A. Karimpour-Ghahnavieh, R. Piergentili, and M.G. Giansanti. 2020. Oncogenic roles of GOLPH3 in the physiopathology of cancer. *International Journal of Molecular Sciences*. 21. doi:10.3390/ijms21030933.

Sharpe, H.J., T.J. Stevens, and S. Munro. 2010. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell*. 142:158–169. doi:10.1016/j.cell.2010.05.037.

Stowell, S.R., T. Ju, and R.D. Cummings. 2015. Protein glycosylation in cancer. *Annual Review of Pathology: Mechanisms of Disease*. 10:473–510. doi:10.1146/annurev-pathol-012414-040438.

Sun, X., B. Chen, Z. Song, and L. Lu. 2021. A quantitative study of the Golgi retention of glycosyltransferases. *bioRxiv*. 2021.02.15.431224. doi:10.1101/2021.02.15.431224.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021 The UniProt Consortium. *Nucleic Acids Research*. 49. doi:10.1093/nar/gkaa1100.

Tran, D.T., and K.G. Ten Hagen. 2013. Mucin-type o-glycosylation during development. *Journal of Biological Chemistry*. 288:6921–6929. doi:10.1074/jbc.R112.418558.

Tu, L., and D.K. Banfield. 2010. Localization of Golgi-resident glycosyltransferases. *Cell Mol Life Sci*. 67:29–41. doi:10.1007/s00018-009-0126-z.

Tu, L., W.C. Tai, L. Chen, and D.K. Banfield. 2008. Signal-mediated dynamic retention of glycosyltransferases in the Golgi. *Science*. 321:404–407. doi:10.1126/science.1159411.

Vajaria, B.N., and P.S. Patel. 2017. Glycosylation: a hallmark of cancer? *Glycoconjugate Journal*. 34:147–156. doi:10.1007/s10719-016-9755-2.

Varki, A., R.D. Cummings, M. Aebi, N.H. Packer, P.H. Seeberger, J.D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J.J. Prestegard, R.L. Schnaar, H.H. Freeze, J.D. Marth, C.R. Bertozzi, M.E. Etzler, M. Frank, J.F.G. Vliegenthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K.F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, and S. Kornfeld. 2015. Symbol nomenclature for graphical representations of glycans. *Glycobiology*. 25:1323–1324. doi:10.1093/glycob/cwv091.

Welch, L.G., and S. Munro. 2019. A tale of short tails, through thick and thin: investigating the sorting mechanisms of Golgi enzymes. *FEBS Letters*. 593:2452–2465. doi:10.1002/1873-3468.13553.

Wood, C.S., K.R. Schmitz, N.J. Bessman, T.G. Setty, K.M. Ferguson, and C.G. Burd. 2009. PtdIns4P recognition by Vps74/GOLPH3 links PtdIns 4-kinase signaling to retrograde Golgi trafficking. *J Cell Biol*. 187:967–975. doi:10.1083/jcb.200909063.

Figure Legends

Figure 1. GOLPH3 and GOLPH3L interact with the COPI coat and a host of Golgi-resident vesicular cargo proteins.

(A) Volcano plots comparing spectral intensity values generated from GST pulldowns from HEK293T cell lysate using GST-tagged GOLPH3 or GOLPH3L vs GST alone. -log P values were generated from Welch's t-tests. Indicated are Golgi-resident integral membrane proteins (Swiss-Prot database, magenta) and COPI coat subunits (green). Data was from three independent biological replicates analysed using Perseus. **(B)** Immunoblot of GST pull-downs as in (A), β-COP (COPI coat subunit) and GALNT7 as a representative cargo. N=2.

Figure 2. The cytoplasmic tails of GALNT2 and ST6GAL1 are sufficient to bestow Golgi retention in a GOLPH3+3L-dependent manner.

(A) The GFP-tagged type II transmembrane reporters for Golgi retention. The cytoplasmic tail of the plasma membrane reporter sucrase-isomaltase (SI) was substituted for that of GALNT2 (novel client) or ST6GAL1 (previously reported client). **(B)** Immunoblots of whole cell lysate from wild-type (WT) U2OS cells and a *ΔΔGOLPH3, GOLPH3L* (ΔΔ) U2OS cell line generated by CRISPR-Cas9 gene-editing. **(C)** Confocal micrographs of stable cell lines expressing the indicated GFP-tagged reporters in a wild-type or GOLPH3 family knockout background. Cells are labelled for golgin-84 as Golgi marker and with a GFP booster. Scale bars, 10 μm.

Figure 3. A quantitative Golgi retention assay to interrogate sorting signals and the contribution of the GOLPH3 proteins.

(A) A schematic of the *in vivo* Golgi retention assay. Under conditions of Golgi retention, the reporter is sequestered in intracellular compartments (primarily the Golgi and COPI vesicles) and the lumenal GFP-FLAG tag is inaccessible to an A647-conjugated anti-GFP antibody under non-permeabilising conditions. In contrast, reporters that are not retained can reach the plasma membrane where the GFP-FLAG tag becomes accessible to the conjugated antibody. The A647 and GFP signals are then analysed by flow cytometry. **(B)** Illustrative flow cytometry data for the Golgi retention assay. Overlaid scatter plots (left) of U2OS cells expressing different chimeric reporters in different genetic backgrounds (above) and corresponding histograms displaying the A647:GFP values. Scatter plots and histograms represent 10,000-20,000 events, N=4. Gating strategy is shown in Fig. S2. **(C** and **D)**, as in (B) but histograms represent 500 and 10,000 - 20,000 events, n=3 and n=1 respectively (wild-type (WT), *ΔΔGOLPH3, GOLPH3L* (ΔΔ)).

Figure 4. Deletion of *GOLPH3* and *GOLPH3L* results in the destabilisation of a diverse array of Golgi-resident enzymes.

(A) Instability of Golgi-resident cargoes (GALNT7 and GPP130) upon the deletion of both GOLPH3 genes. Immunoblots of whole cell lysates from wild-type, *ΔGOLPH3;ΔGOLPH3L* and polyclonal rescue U2OS cells. **(B)** Volcano plot comparing spectral intensity values for individual proteins in

ΔGOLPH3;ΔGOLPH3L U2OS cells vs *ΔGOLPH3;ΔGOLPH3L +GOLPH3* polyclonal rescue U2OS cells. The data set was generated from two repeats, was Z-score normalised according to the median and -log P values were generated from a Student's t-tests. COPI subunits (green), Golgi-resident integral membrane proteins (magenta), all other glycoproteins (cyan), based on Swiss-Prot. **(C)** Data from (A) for relative protein abundances in *ΔGOLPH3;ΔGOLPH3L* cells vs *ΔGOLPH3;ΔGOLPH3L +GOLPH3* rescue cells plotted against the data for GST-GOLPH3/3L binding vs GST binding from Table S2 (combined data set in Table S3C). COPI coat and GOLPH3 proteins (green) and Golgi-resident integral membrane proteins (magenta), and the dotted line shows the cut-off for degradation, below which all proteins are Golgi residents. **(D)** A table of the highest confidence GOLPH3+3L interactors named in (C), and defined by showing greater degradation than any non-Golgi protein. All are type II, apart from Glg1 (type I), PGAP4 (3 TMDs) and EBAG9 (unclear). **(E)** Flow cytometry of lectin binding to *ΔGOLPH3;ΔGOLPH3L* and wild-type U2OS cells. FITC-conjugated lectins with different specificities (lectins recognising O-linked glycans: VVA and jacalin). Lectin specificity was validated using saturating concentrations of the indicated competing sugar. Histograms are normalised to the mode value for each treatment. At least 10,000 events were collected for each cell line (Fig. S2), and the plots shown are representative of three biological replicates. Symbol nomenclature for glycans was used for illustrations (Varki et al., 2015).

Figure 5. GOLPH3 interacts with the short, positively-charged, cytoplasmic tails of a wide range of type II transmembrane proteins of the Golgi.

(A) A schematic showing the Golgi enzyme cytoplasmic tail chimeras used in binding experiments. **(B and C)** Tests of the ability of GST-GOLPH3 to pulldown different cytoplasmic tail chimeras from HEK293T cell lysate. Tail sequences (in brackets) are defined according to Uniprot. Charged residues are coloured in blue (positive) or red (negative). The predicted total charges of the tails are based on a cytosolic pH of 7.4 and include the positive charge of the amino terminus. The tail of sucrase-isomaltase (SI) is predicted to be phosphorylated at the serine at position 7 (underlined). Bold letters indicate changes resulting from targeted mutagenesis. All blots were imaged by chemiluminescence unless otherwise stated (Alexa Fluor 555). Data representative of two independent replicates. **(D)** Histograms displaying the A647:GFP values from an *in vivo* Golgi retention assay comparing U2OS cells expressing the SI reporter with the membrane-proximal insertion of polybasic stretches in different genetic backgrounds (wild-type (WT) or *ΔGOLPH3;ΔGOLPH3L* (ΔΔ). Data collected in the same experiment as Fig. 3 C, and so shares the same SI reporter/wild-type (WT) control. See Fig. S2 for gating strategies. Histograms correspond to 500 events and are representative of 3 independent replicates. **(E)** Confocal micrographs of the indicated GFP-tagged reporters stably expressed in U2OS cells. Cells are labelled with a GFP-booster and for golgin-84 (Golgi marker). Scale bars, 10 μm.

Figure 6. The clientele of GOLPH3 and GOLPH3L have an enrichment of positively-charged residues in the membrane-proximal region of their cytoplasmic tails.

(A) Transmembrane span regions of type II membrane proteins from different sets of proteomic analyses, showing differences in relative positional abundances of amino acids. Sequences were

aligned according cytoplasmic TM span edge (position 15) and plots use "X" to represent the absence of an amino acid; for positions beyond the sequence edge. The two upper panels are derived from GST-GOLPH3 pulldowns and a reported COPI proteome (Figure 5A-C, proteins with a mean log2 SILAC ratio of > 0, sample vs control; Adolf et al., 2019): of the type II proteins present in the COPI proteome, those enriched in GST pulldowns by GOLPH3+3L were compared to those that were not enriched. The two lower panels are type II proteins that showed the largest reduction in *ΔGOLPH3;ΔGOLPH3L* cells vs *ΔGOLPH3;ΔGOLPH3L +GOLPH3* rescue cells compared to all other type II proteins in the dataset (Fig. 4 B, Log2[*ΔGOLPH3;ΔGOLPH3L*/GOLPH3 rescue] values: degraded ≤ -0.1, non-degraded > -0.1). Proteins listed in Table S4. **(B)** Quantification of membrane-proximal positively-charged residues in the cytoplasmic tails of type II proteins from the endoplasmic reticulum (ER), Golgi and plasma membrane (PM, Uniprot). Positive charges were counted within the 6 membrane-proximal residues; arginine or lysine (at pH 7.4) and the α-amino group for tails ≤ 6 residues long. **(C)** as in (B) but comparing cytoplasmic tail length. **(D)** Categorisation of degraded and non-degraded proteins according to their position in various glycosylation pathways. Proteins were assigned to different glycosylation pathways according to Schjoldager et al., 2020. *N=155 rather than 154 since FUT8 can function in both capping and core extension. **(E)** A model for GOLPH3 membrane and cargo docking. The structure of GOLPH3 reveals a flat surface containing a large electronegative patch (red, negative; blue, positive, (Wood et al., 2009). This electronegative patch could interact with the positively-charged residues of the cytoplasmic N-termini of clients, and thus recruit them into COPI-coated vesicles.

Supplemental Figure Legends

Figure S1. Validation of the *ΔGOLPH3;ΔGOLPH3L* U2OS cell line.

(A) Agarose gels resolving PCR-amplified regions of the genomic loci of *GOLPH3* and *GOLPH3L* targeted by CRISPR-Cas9 gene-editing. Base pair (bp). **(B)** *GOLPH3* was targeted at one site in exon 2 (plasmid pLGW443 encoding guide GAGAGGAAGGTTACAACTAG (green line)), inducing small indel mutations in at least two alleles (lower band) and a 197 bp out-of-frame insertion in another allele (upper band). **(C)** *GOLPH3L* was targeted at a site in exon 2 and a site in intron 2-3 (plasmids pLGW444 for guide CTTCTTCCATAAGAGTAAGG and pLGW445 for GTAATGCAGTTAGGTTTGCT), inducing a 62 bp deletion in one allele and a 79 bp deletion with a 234 bp insertion in the other allele. **(D)** A volcano plot comparing spectral intensity values corresponding to the individual proteins in *ΔGOLPH3;ΔGOLPH3L* cells vs wild-type U2OS cells. The data sets were generated from a duplicate of repeats, were Z-score normalised according to the median and -log P values generated with a Student's t-test. Points correspond to individual proteins. Notable proteins displaying a large difference are coloured: GOLPH3 proteins and COPI (green), Golgi-resident cargo (magenta).

Figure S2. Flow cytometry gating strategies.

A representative gating strategy for the *in vivo* Golgi retention assays and the lectin staining. Wild-type stained U2OS cells were used to assign the gates. Plots represent 25,000 events. Hierarchical gating strategy in order from 1-4: top panels show isolation of singlets based on forward (FSC) and side (SSC) scatter using height (H) and width (W). Lower panels show gating for live cells using an eFluor 780 fixable live/dead stain followed by a gate for GFP-positive cells (or FITC positive cells for lectin stains). Compensation was done using single colour controls on an LSRII, and plots and gates were generated using FlowJo V10.

Figure S3. Genetic rescue of the destablisation of Golgi residents in *ΔΔGOLPH3;ΔGOLPH3L* cells.

(A) Confocal micrographs of wild-type and *ΔGOLPH3;ΔGOLPH3L* U2OS cells or a polyclonal rescue cell line stably re-expressing *GOLPH3* in the *ΔGOLPH3;ΔGOLPH3L* background. Arrows indicate cells which do not express detectable levels of GOLPH3 and so lack the rescue of the loss of the Golgi resident GALNT2. TGN46 (Golgi/TGN marker). **(B)** As in (A) but wild-type or *ΔGOLPH3;ΔGOLPH3L* U2OS cell lines transiently transfected with plasmids encoding GOLPH3 or GOLPH3L as indicated, and labelled for GALNT7 (Golgi resident enzyme) and GM130 (Golgi marker). Scale bars, 10 μm.

Figure S4. GOLPH3 interacts with the tails of GALNT2 and GALNT7.

Test of the ability of bacterially expressed GST-tagged GOLPH3 to pulldown different GFP-FLAG-tagged cytoplasmic tail chimeras from HEK293T cell lysate. Experiment representative of two independent replicates. Sequences of the cytoplasmic tails are given below, charged residues are charged are coloured blue (positive) or red (negative). Note that the tail of sucrase-isomaltase (SI) is predicted to be phosphorylated on the serine at position 7.

Supplemental Tables

Table	Associated Figure(s)	Description	Sheets
S1	Fig. 1 A	Affinity chromatography mass spectrometry: GST-GOLPH3 vs GST, GST-GOLPH3L vs GST	A: Raw mass spectrometry data, Welch's t-test B: Data as plotted in Fig. 1 A, filtered for COPI proteins (green) and Swiss-Prot Golgi integral membrane proteins (magenta). C: Swiss-Prot Golgi integral membrane proteins reference list
S2	N/A	Affinity chromatography mass spectrometry: GST-GOLPH3 and GST-GOLPH3 combined vs GST, compared to reference COPI proteome	A: Raw mass spectrometry data, Welch's t-test B: GOLPH3+3L interactome (proteins enriched in GST-GOLPH3+3L sample with P ≤ 0.05, Welch's t-test) compared to a COPI Proteome (Figure 5A-C, proteins with a mean log2 SILAC ratio of > 0, sample vs control; Adolf et al., 2019)
S3	Fig. 4 B,C; Fig S1 D	TMT-labelled comparative proteomics: Wild-type vs *ΔGOLPH3;ΔGOLPH3L* vs *ΔGOLPH3;ΔGOLPH3L +GOLPH3* (polyclonal)	A: Raw mass spectrometry data, Z-score normalised to median. B: *ΔΔGOLPH3*, *GOLPH3L* vs *ΔΔGOLPH3*, *GOLPH3L +GOLPH3* polyclonal rescue (Z-score normalised), Student's t-test, filtered for COPI proteins (green), Golgi integral membrane proteins (magenta) and other glycoproteins (cyan) according to Swiss-Prot. As plotted in Fig. 4 B. C: Wild-type vs *ΔΔGOLPH3*, *GOLPH3L* (Z-score normalised), Student's t-test, filtered for COPI proteins (green) and Golgi-resident cargo proteins of interest (magenta). As plotted in Fig. S1 D. D: *ΔΔGOLPH3*, *GOLPH3L* vs *ΔΔGOLPH3*, *GOLPH3L +GOLPH3* polyclonal rescue protein abundance log2 difference plotted against GST-GOLPH3+3L vs GST affinity chromatography log2 difference. As plotted in Fig. 4 C. E: Swiss-Prot glycoprotein reference list (excluding nuclear and cytoplasmic proteins)
S4	Fig. 6	Proteins used in bioinformatic analyses	A: COPI+GOLPH3+: type II proteins present in both the COPI proteome and GOLPH3+3L interactome. COPI+GOLPH3-: type II proteins present in the COPI proteome but **not** the GOLPH3+3L interactome. As analysed in Fig. 6 A (derived from Table S2) B: Degraded: type II proteins that were at the lowest levels in *ΔΔGOLPH3, GOLPH3L* cells vs *ΔΔGOLPH3, GOLPH3L +GOLPH3* rescue cells. Non-Degraded: all other type II proteins in the dataset, derived from Table S3 B, as analysed in Fig. 6 A (Log2[ΔΔGOLPH3, GOLPH3L/GOLPH3 rescue] values: Degraded ≤ -0.1, Non-Degraded > -0.1). Degraded and Non-Degraded groups categorised according to function as plotted in Fig. 6 D. C: ER, Golgi and plasma membrane-resident type II proteins according to Uniprot. As analysed in Fig. 6 B,C.
S5	N/A	Plasmids and Antibodies	A: Plasmid list B: Primary antibody list C: Secondary antibody list

Figure 1

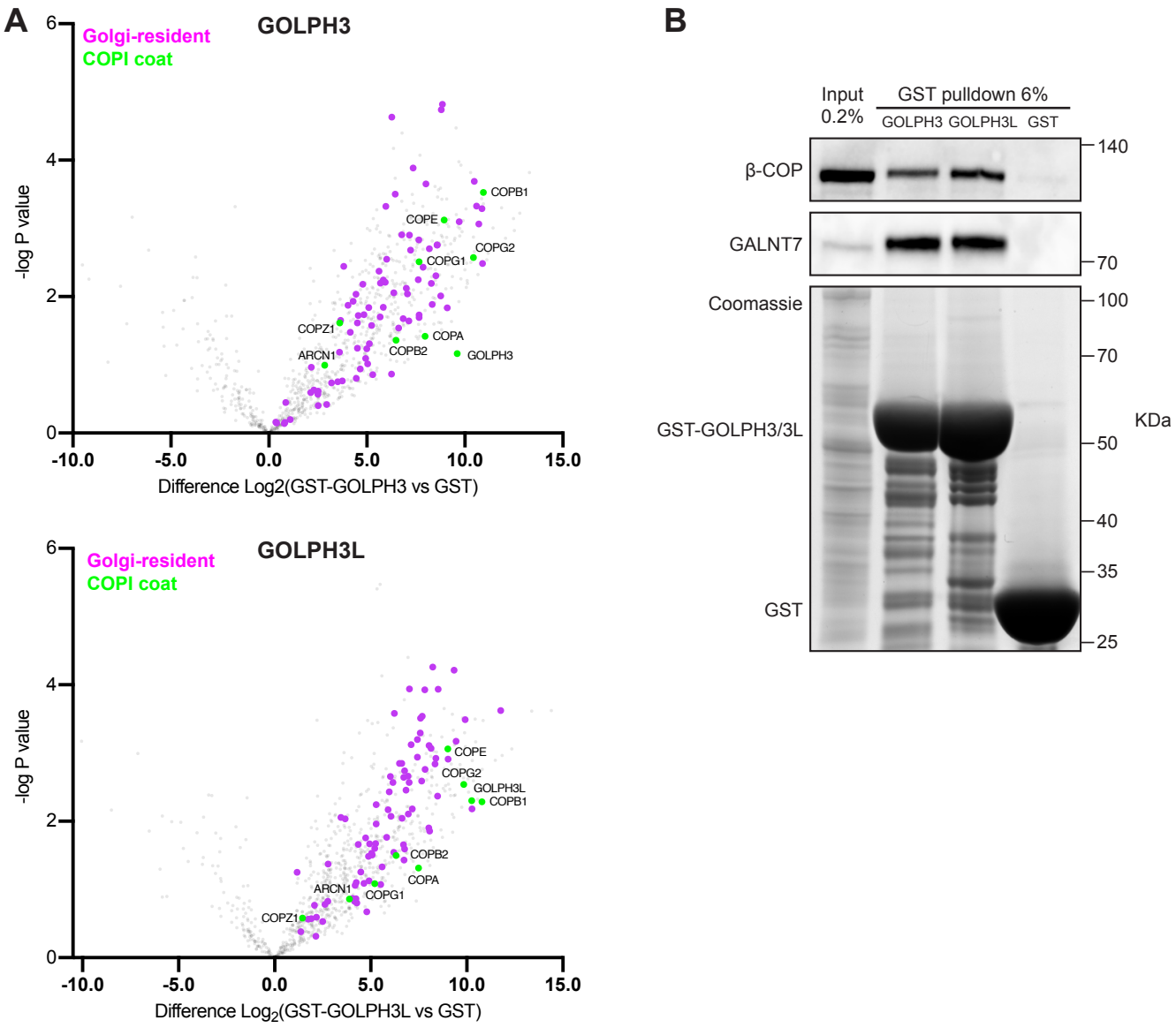


Figure 2

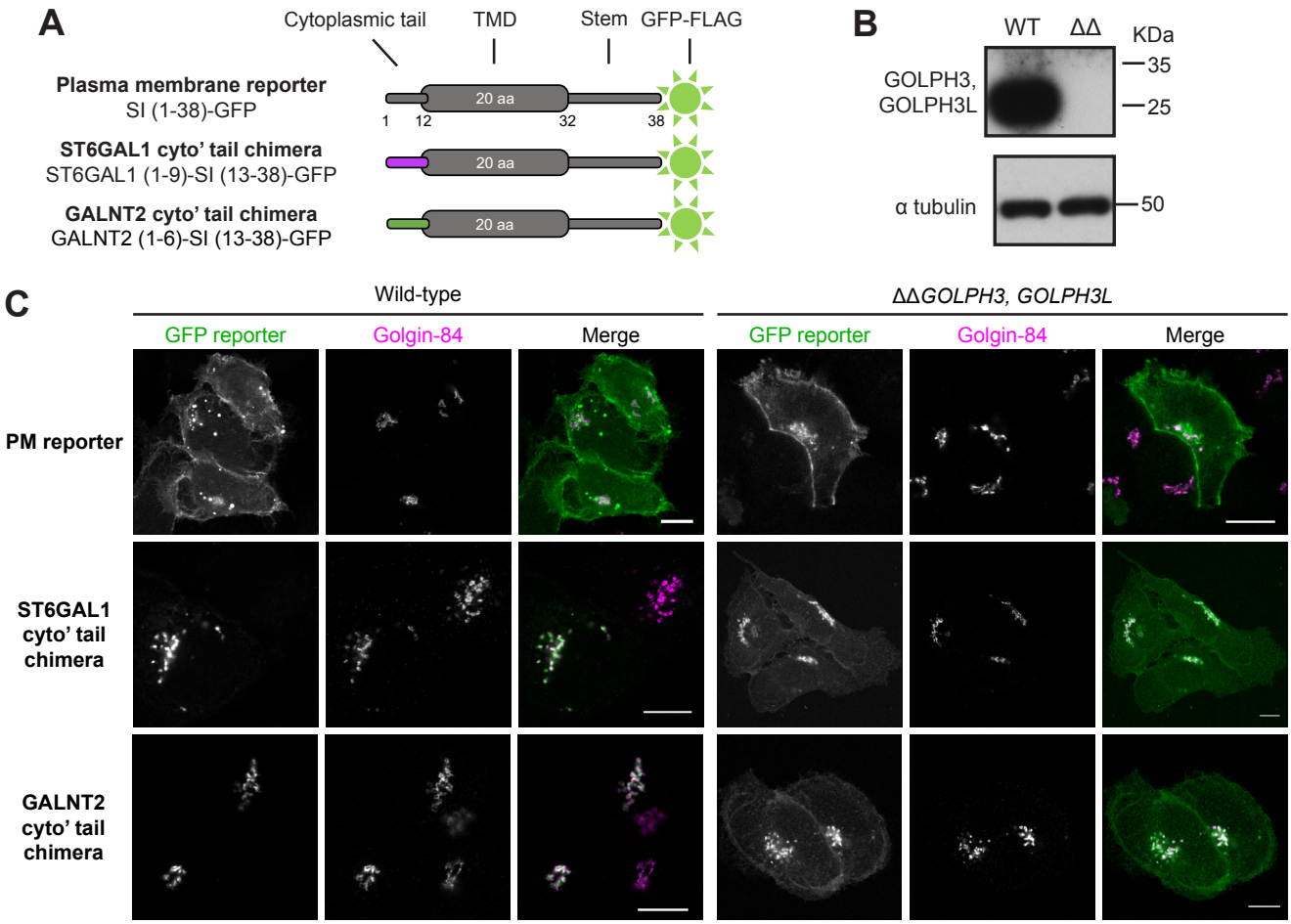


Figure 3

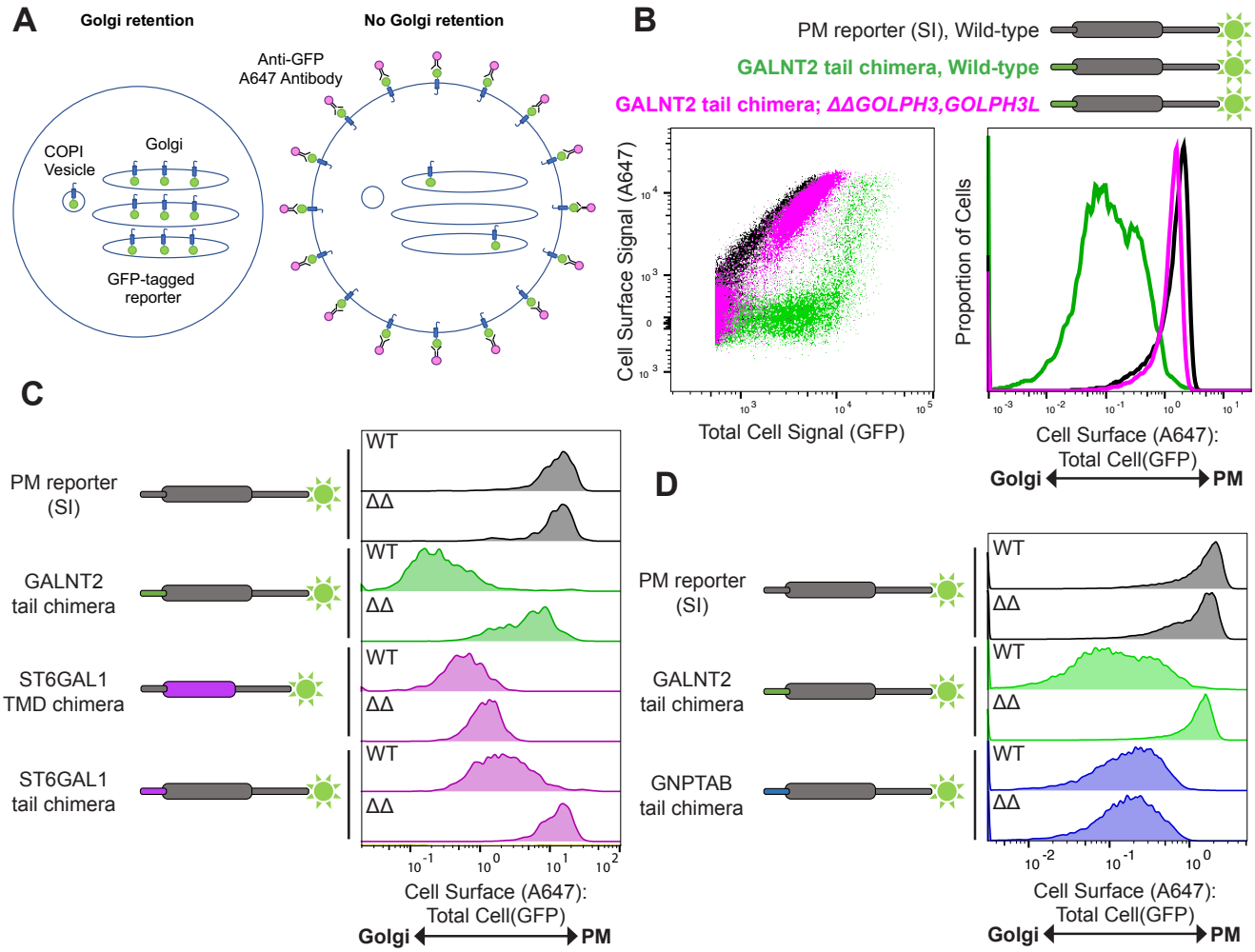
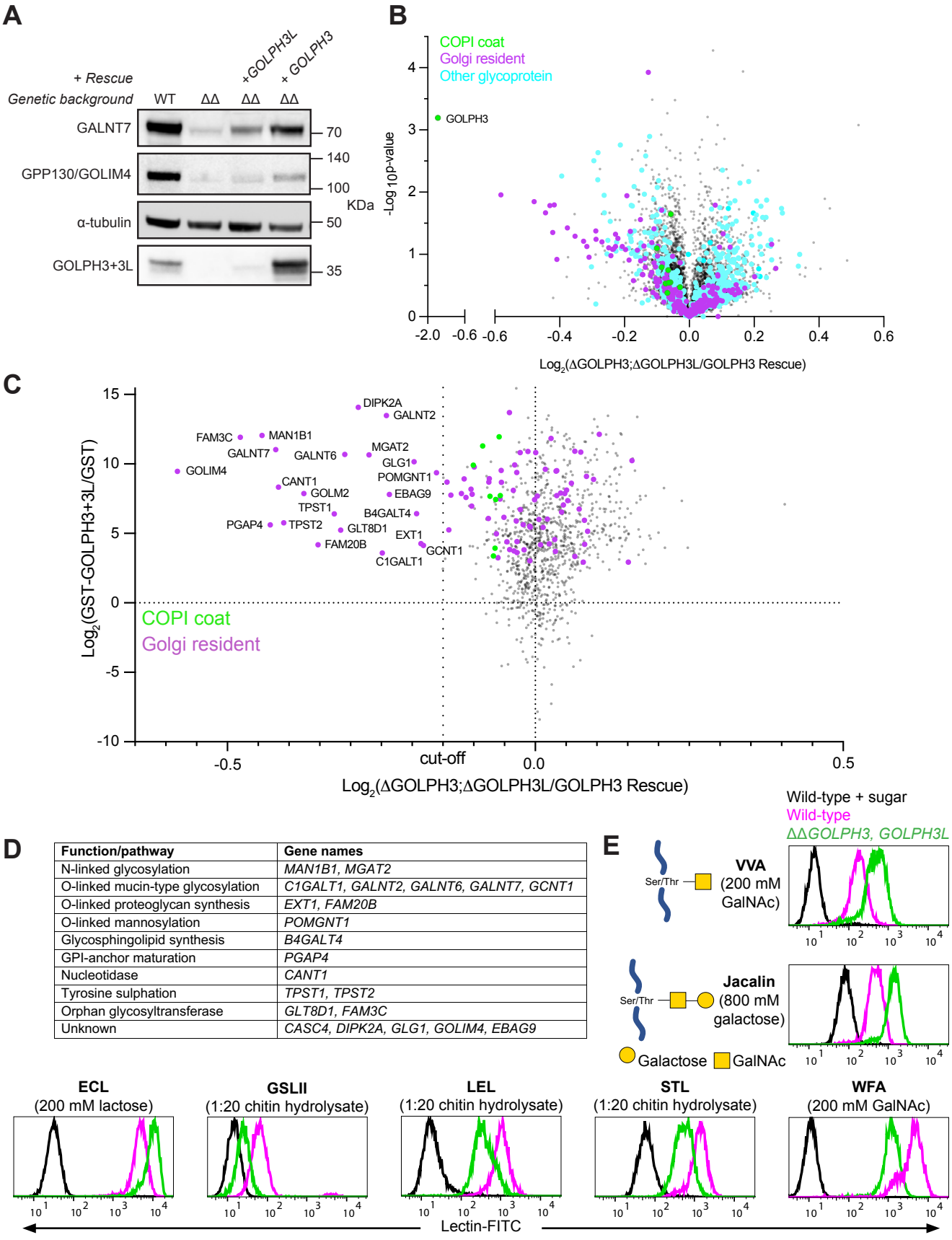


Figure 4



Function/pathway	Gene names
N-linked glycosylation	*MAN1B1, MGAT2*
O-linked mucin-type glycosylation	*C1GALT1, GALNT2, GALNT6, GALNT7, GCNT1*
O-linked proteoglycan synthesis	*EXT1, FAM20B*
O-linked mannosylation	*POMGNT1*
Glycosphingolipid synthesis	*B4GALT4*
GPI-anchor maturation	*PGAP4*
Nucleotidase	*CANT1*
Tyrosine sulphation	*TPST1, TPST2*
Orphan glycosyltransferase	*GLT8D1, FAM3C*
Unknown	*CASC4, DIPK2A, GLG1, GOLIM4, EBAG9*

Figure 5

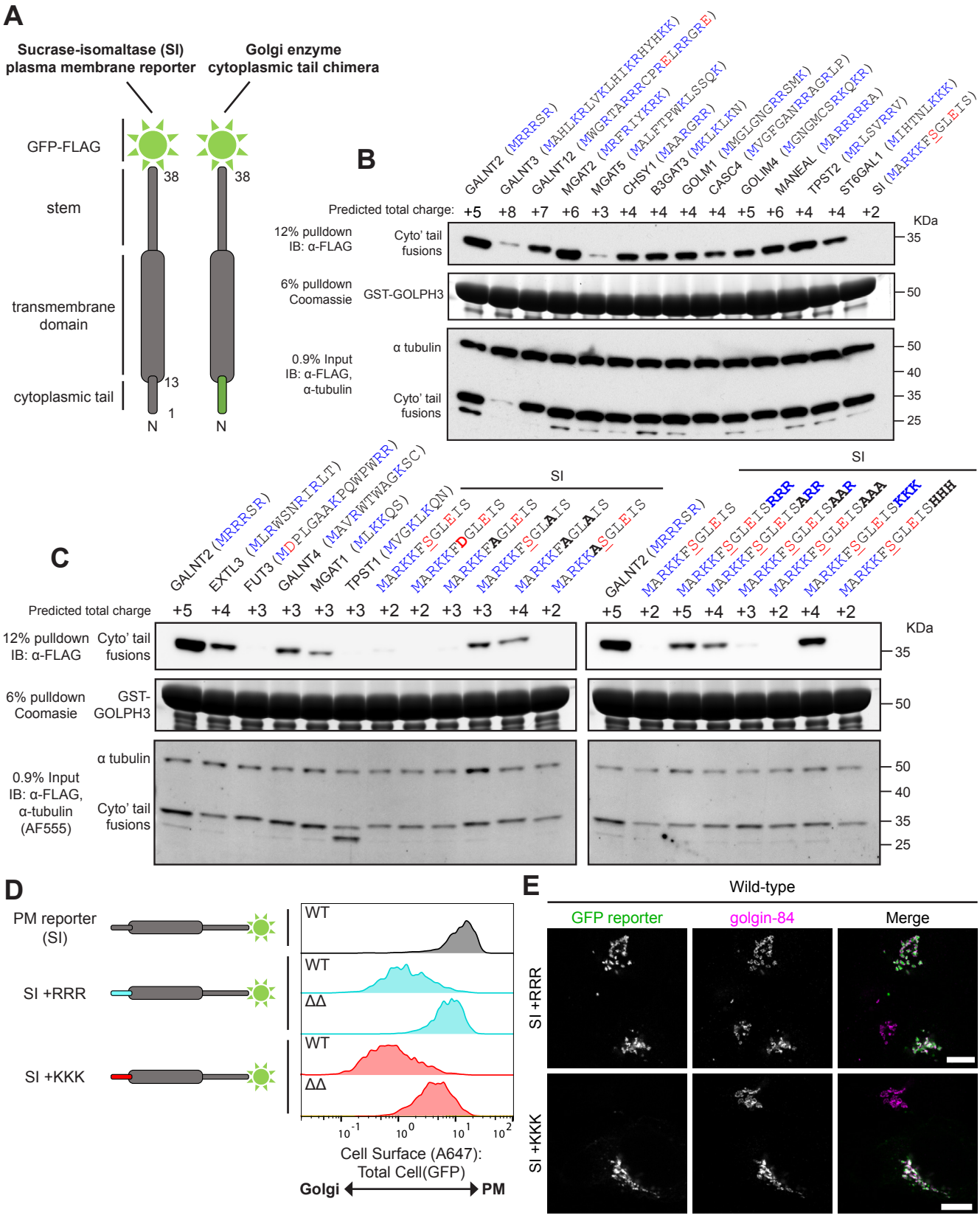


Figure 6

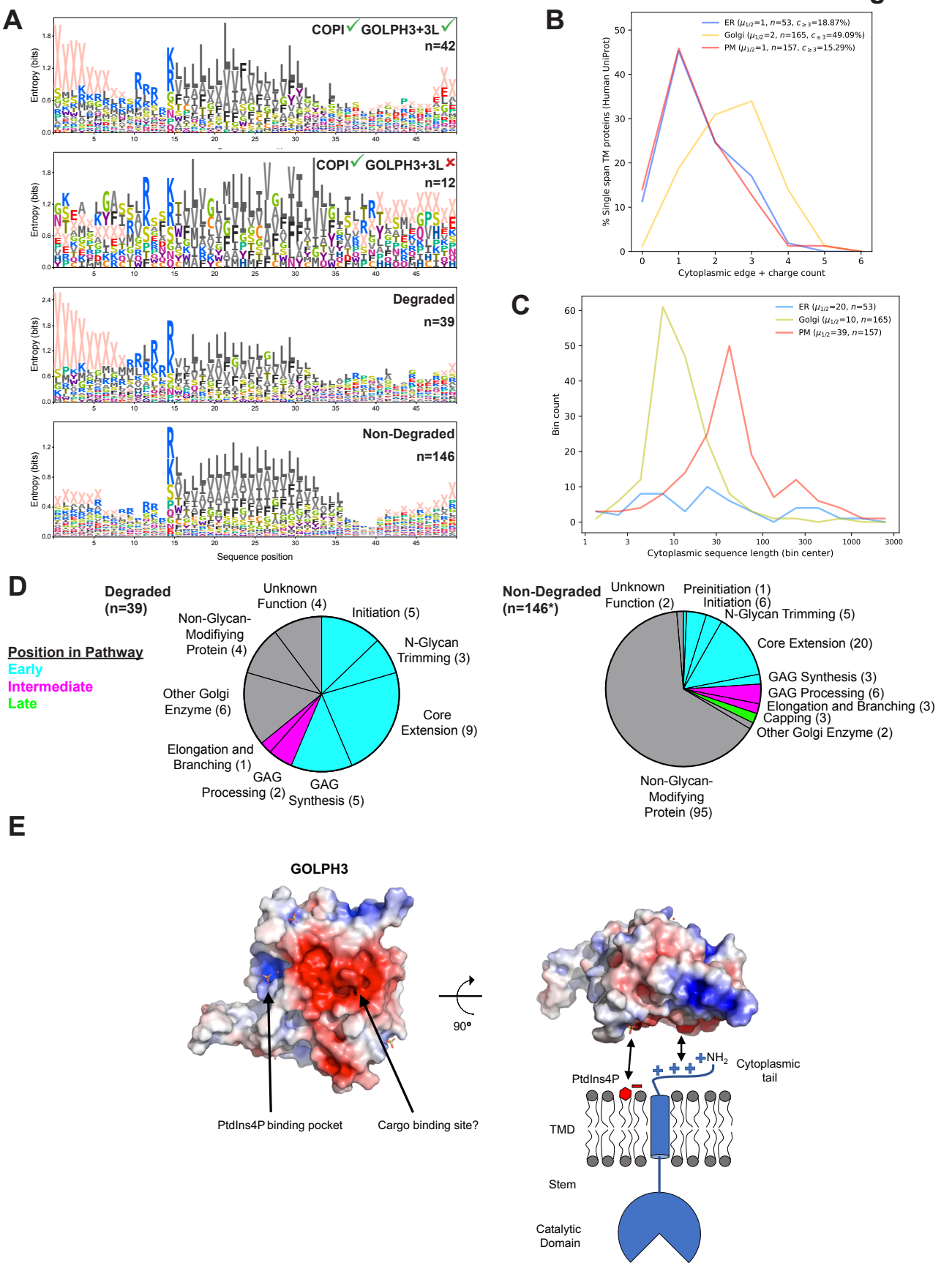


Figure S1

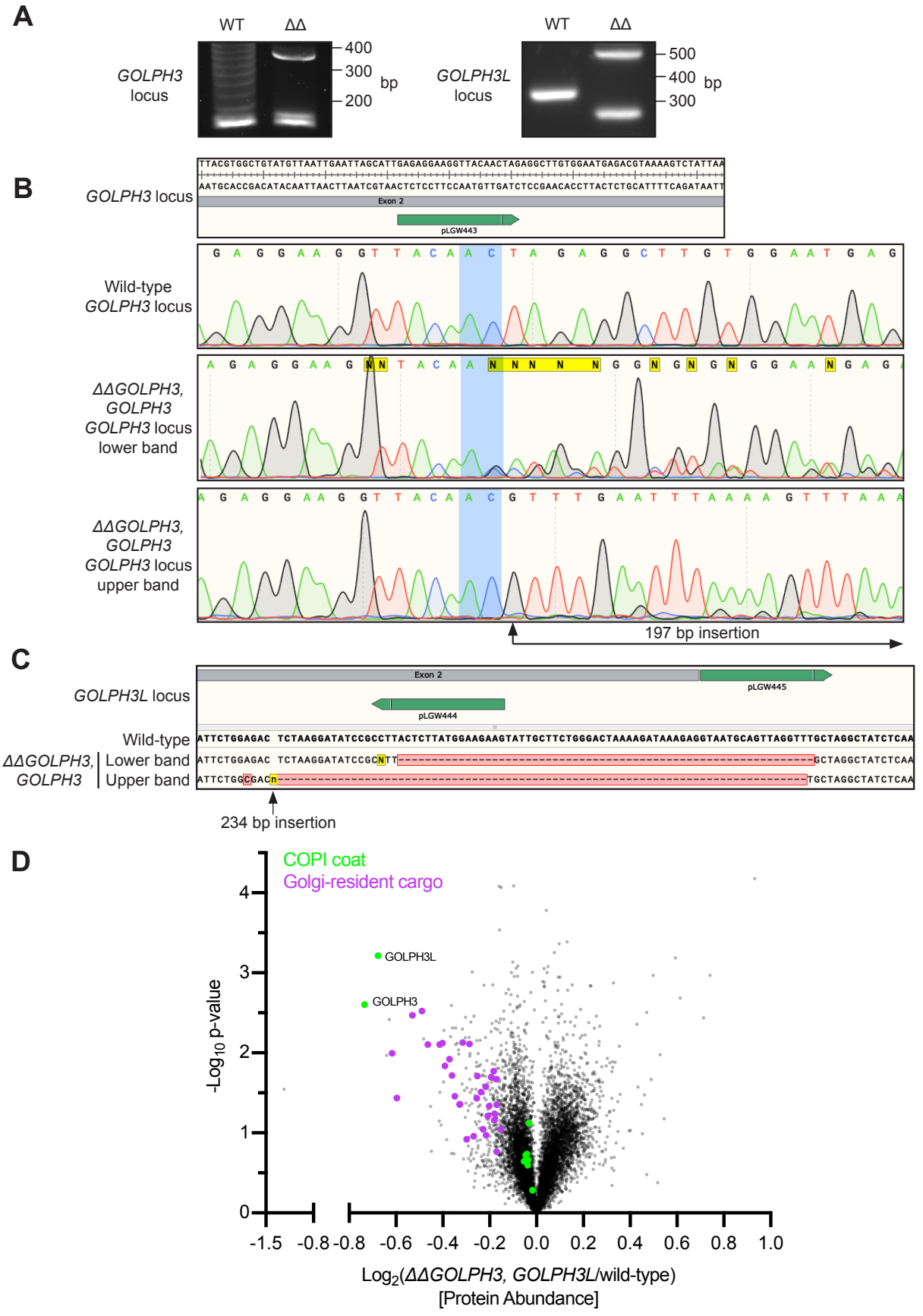


Figure S2

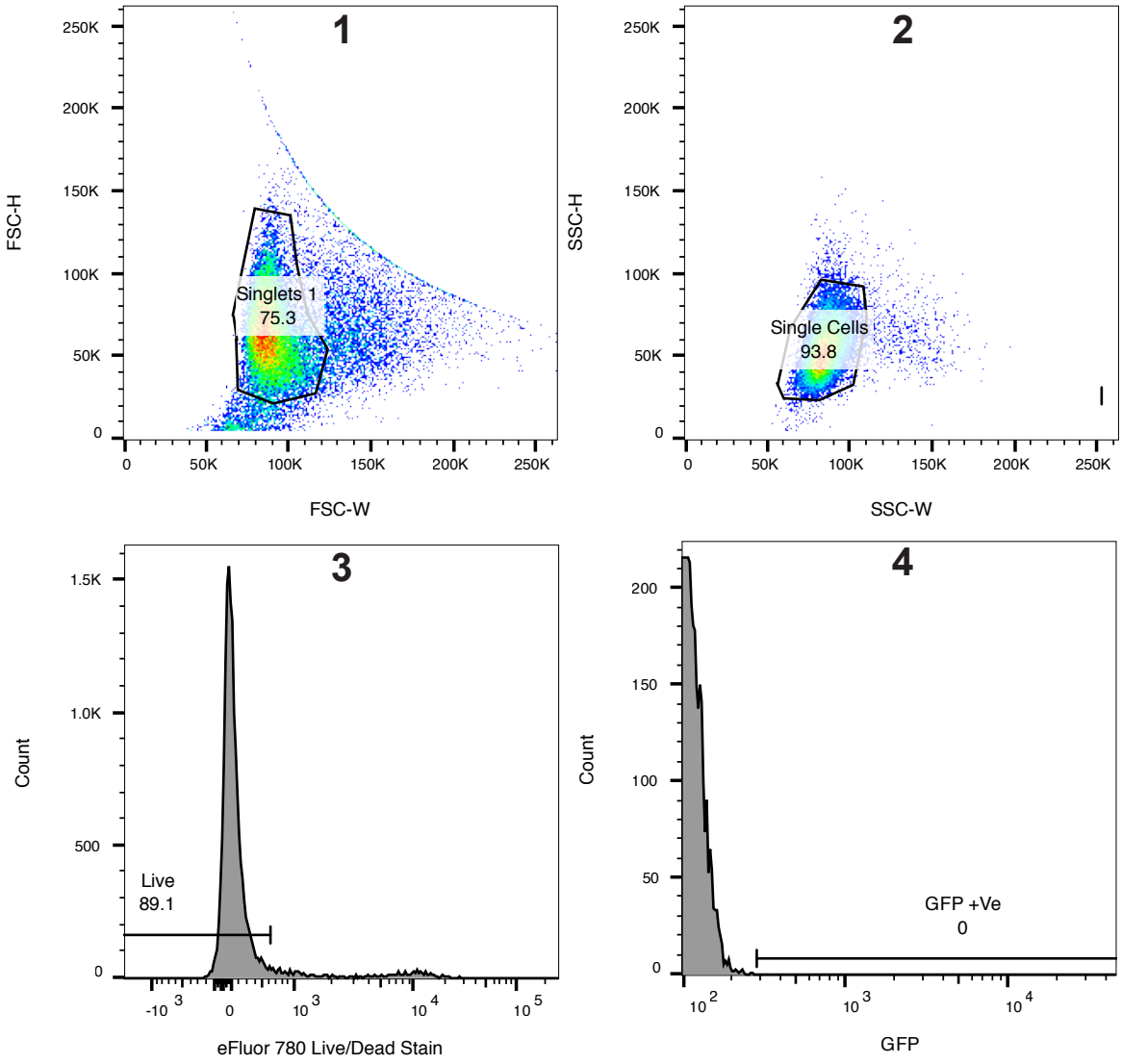


Figure S3

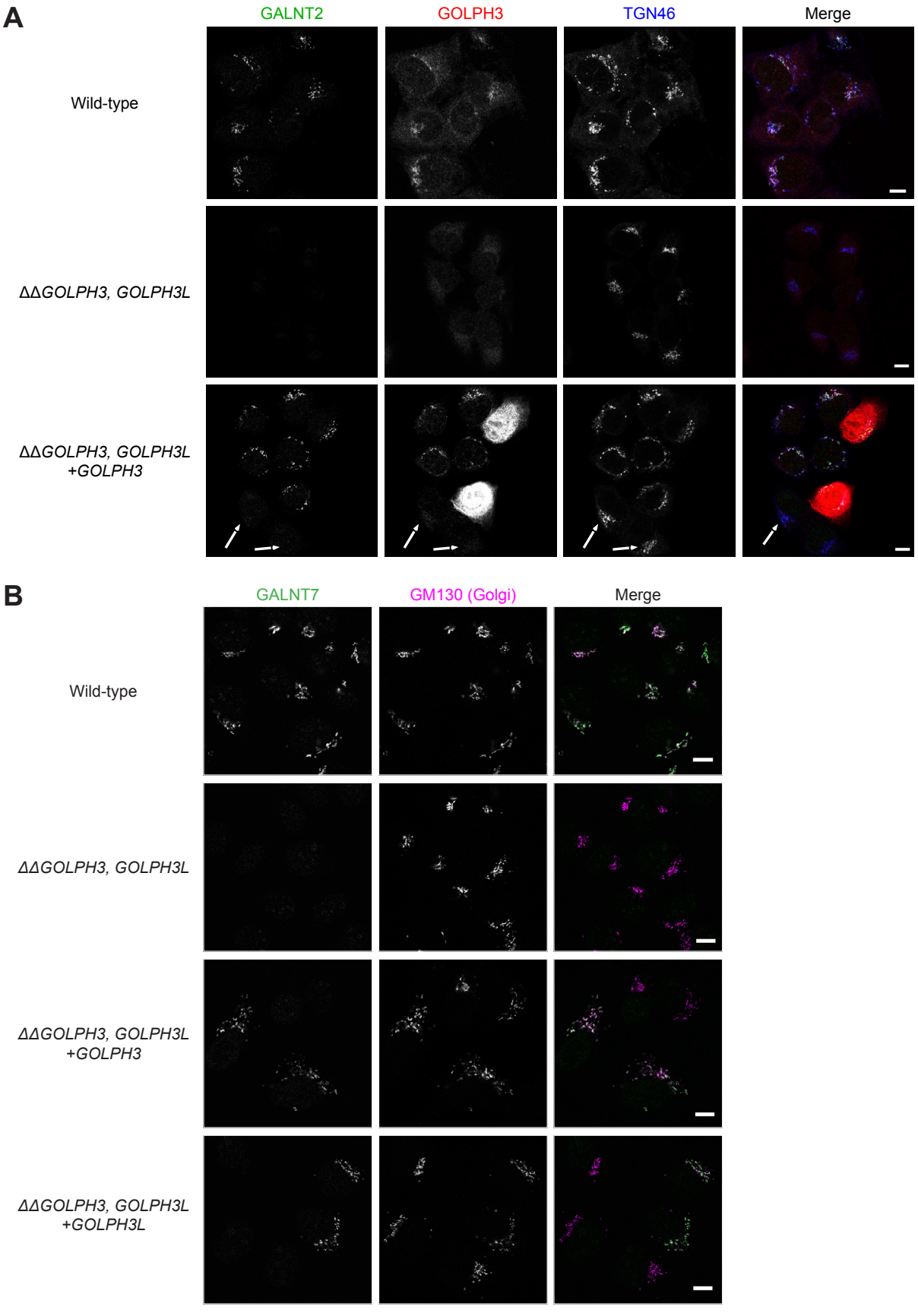


Figure S4