Investigating the trafficking of Golgi-resident glycosylation enzymes

Lawrence George Welch

MRC Laboratory of Molecular Biology
Pembroke College
University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

Under the supervision of Dr. Sean Munro
March 2020

Declaration

This thesis is submitted for the degree of Doctor of Philosophy. This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed 60,000-word limit for the Faculty of Biology Degree Committee.

Lawrence George Welch

28th March 2020

Summary

Name - Lawrence George Welch
Title - Investigating the trafficking of Golgi-resident glycosylation enzymes

The Golgi is a major site of glycan processing mediated by a large number of Golgi-resident glycosylation enzymes. These enzymes are largely type II transmembrane domain (TMD) proteins consisting of a short N-terminal cytosolic tail, a short TMD and a lumenal 'stem' region which separates the catalytic domain and the lipid bilayer. The cytosolic tail, TMD, and stem together make the CTS domain which is responsible for the Golgi-targeting of these enzymes via multiple COPI-dependent mechanisms. What is not clear is the relative contribution of each mechanism to Golgi retention or how they work to fine tune the sub-Golgi localisation of these enzymes. To explore these nuances of intra-Golgi traffic, a MitoID proteomic screen was applied to the intra-Golgi golgin tethers GMAP-210 and golgin-84 to identify novel intra-Golgi vesicle-resident factors. The screen identified the Rab2 effector of unknown function FAM114A2. FAM114A2 and paralogous protein FAM114A1 were shown to specifically associate with early-Golgi membranes. A predicted ALPS motif was necessary for membrane association, presumably through non-specifically absorbing onto highly-curved membranes. Other regions of the proteins were required to mediate a specific association with membranes of the early-Golgi. GST affinity chromatography and mass spectrometry revealed that FAM114A1 pulled down Rab2A and Rab2B while FAM114A2 pulled down several COPI cargo proteins including many glycosylation enzymes. FAM114A2 interacted directly with the cytoplasmic tails of cargo which had a membrane-proximal polybasic motif. Similar binding studies revealed that the COPI adaptors GOLPH3+3L pulled down the COPI coat and over a hundred novel clients. GOLPH3+3L were shown to interact promiscuously with the cytoplasmic tails of clients with a total predicted charge of ≥4+. An in vivo Golgi retention assay revealed that tails which satisfy the GOLPH3+3L charge threshold are sufficient to impart Golgi-retention in a GOLPH3+3L-dependent manner. The deletion of FAM114A genes had no impact on retention in this assay. The assay also showed that the TMD of ST6GAL1 was sufficient to impart Golgi retention in a GOLPH3+3L-independent manner. The deletion of GOLPH3+3L, but not FAM114A genes, triggered the lysosomal degradation of COPI cargo and caused a global defect in glycosylation in U2OS cells. In contrast, deletion of the D. melanogaster FAM114A ortholog CG9590 caused a defect in mucin-type glycosylation in adult flies. CG9590 interacted directly with the COPI coat and pulled-down a host of COPI cargo proteins including mucin-type glycosylation enzymes. It is likely that FAM114A is an evolutionary-conserved COPI adaptor which acts to recycle a specific subset of COPI cargo from the early-Golgi. In contrast, GOLPH3 proteins are promiscuous COPI adaptors which function as master gatekeepers at the late-Golgi to recycle and retain a broad array of Golgi-residents.

Acknowledgements

I would like to thank my supervisor, Sean Munro, for his guidance and support during the course of my PhD and for the exciting opportunity to work in his group. I would also like to thank him for the great patience he has shown toward my attempt at scientific wordplay and my colourful biological analogies. My thanks also go to my second supervisor Liz Miller and my university supervisor Paul Dupree for their stimulating discussions on my project.

My special thanks go to the whole Munro lab, past and present, for making the lab a truly lovely environment to work in. To work amongst such an easy-going, thoughtful and cohesive group of friends has been a privilege and a blessing. In particular, I would like to thank Alison Gillingham, John Shin, Antonio Galindo and Stephen Barratt for their seemingly endless supply of reagents, technical support and scientific discussion. I would like to thank Tim Stevens for his help in conducting bioinformatic analyses and Jérôme Cattin-Ortolá for his consistently thoughtful and insightful comments on several presentations and manuscripts I have prepared, including this thesis. My thanks go to Nadine Muschalik for conducting the *Drosophila* genetics shown in this thesis, without her expertise a key breakthrough in the project would have remained a fanciful hope. I would like to thank Jessie Bertram and John Kilmartin for their jovial and considerate presence as my bench neighbours. I would also like to thank all of the participants of whiskey club for the hazy exchange of peaty scotch and scientific ideas alike.

I would like to express my gratitude to the support services and core facilities of the LMB, their selfless help was a great enabling and facilitating force which underpinned all of the work of my PhD. In particular, I would like to thank Farida Begum, Sew-Yeu Peak-Chew and Mark Skehel in the MRC LMB Mass Spectrometry Core Facility, Jianguo Shi of the MRC LMB Baculovirus Facility, and Jamie Ingle of instrument services.

My thanks go to my sister Alice for her help proof-reading this thesis and to the rest of my family and friends for their support. I would also like to thank Chiara for her great patience and support.

Contents

Abbreviations

Abbreviation	Full Name
A4GALT	Lactosylceramide 4-alpha-galactosyltransferase
A647	Alexa Fluor 647
aa	Amino acid
ALDH3A2	Aldehyde dehydrogenase family 3 member A2
Alg3	Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase
ALPS	Amphipathic lipid packing sensor
AP	Adaptor protein
Arf	ADP-ribosylation factor
ARFGAP1	ADP-ribosylation factor GTPase-activating protein 1
ATP5	ATP synthase subunit
B3GALT6	Beta-1,3-galactosyltransferase 6
B3GAT3	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3
B4GALT1	Beta-1,4-galactosyltransferase 1
BFP	Blue fluorescent protein
C16orf13	Methyltransferase-like 26
C1GalT	Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase
CARS	Cysteine--tRNA ligase, cytoplasmic
CASC4	Cancer susceptibility candidate gene 4 protein
Cda5	Chitin deacetylase-like 5
CHPF2	Chondroitin sulfate glucuronultransferase
CHSY1	Chondroitin sulfate synthase 1
Cht5	Chitinase 5
Cht7	Chitinase 7
CRISPR	Clustered regularly interspaced short palindromic repeats
CTS	Domain consisting of the cytoplasmic tail, TMD and stem
DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit
DMEM	Dulbecco’s modified Eagle’s medium
DNA	Deoxyribose nucleic acid
DSL	Datura stramonium lectin
DUF719	Domain of unknow function 719
ECL	Erythrina cristagalli lectin
EDTA	Ethylenediaminetetraacetic acid
EFF1	Elongation Factor 1
EM	Electron microscopy
ER	Endoplasmic reticulum
ERES	ER exit sites

ERGIC	ER-Golgi intermediate compartment
EXT1	Exotosin 1
EXT2	Exotosin 2
EXTL3	Exostosin-like 3
FACS	Fluorescence-activated cell sorting
FAM20B	Glycosaminoglycan xylosylkinase
Fbp1	Fat body protein 1
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
FucTB	Alpha-(1,3)-fucosyltransferase B
FUT3	Galactoside 3(4)-L-fucosyltransferase
GAG	Glycosaminoglycan
GALA	GalNAc-T activation
GalNAc	N-Acetylgalactosamine
GALNT	Polypeptide N-acetylgalactosaminyltransferase (mammalian)
GCC185	185 kDa Golgi coiled-coil protein
GCNT1	Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase
GFP	Green fluorescent protein
GGA	Golgi-localized, γ-ear-containing, ADP ribosylation factor (ARF)-binding proteins
GLG1	Golgi apparatus protein 1
GM130	130 kDa cis-Golgi matrix protein
GMAP-210	Golgi-associated microtubule-binding protein 210
GNB1	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1
GNPTAB	N-acetylglucosamine-1-phosphotransferase subunits alpha/beta
GnTI	N-acetylglucosaminyltransferase I
GO	Gerodermia osteodysplastica
GOLIM4	Golgi integral membrane protein 4
GOLM1	Golgi membrane protein 1
GOLPH3	Golgi phosphoprotein 3
GOLPH3L	Golgi phosphoprotein 3-like
Gp93	Glycoprotein 93
GSLII	Griffonia simplicifolia lectin II
GST	Glutathione S-transferases
GTP	Guanosine-5'-triphosphate
Hexo1	Beta-hexosaminidase
Hmu	Hemomucin
HS2ST1	Heparan sulfate 2-O-sulfotransferase 1
Hts	Protein hu-li tai shao
IPTG	Isopropyl β-D-1-thiogalactopyranoside

IVA	*In vivo* assembly
Jac	Jacalin
LAPTM4B	Lysosome-associated protein transmembrane-4 beta
LEL	Lycopersicon esculentum lectin
LFQ	Label-free quantification
Lsp1	Larval serum protein 1
MANEAL	Glycoprotein endo-alpha-1,2-mannosidase-like protein
MAO	C-terminal transmembrane domain of monoamine oxidase A
Mgat1	Alpha1,3-mannose beta1,2-N-acetylglucosaminyltransferase 1
MGAT2	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
MGAT5	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A
MGST3	Microsomal glutathione S-transferase 3
MHD	μ-homology domain
Mmp1	Matrix metalloproteinase 1
MOGS	Mannosyl-oligosaccharide glucosidase
MRC LMB	Medical Research Council Laboratory of Molecular Biology
NUP133	Nuclear pore complex protein 133
OST	Oligosaccharyltransferase
Oxt	Xylosyltransferase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	0.1% Tween-20 in PBS
PCR	Polymerase chain reaction
PEI	Polyethyleneimine
pgant	Polypeptide N-acetylgalactosaminyltransferase (Drosophila)
PMSF	Phenylmethylsulfonyl fluoride
POMGNT1	Protein O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase 1
PTS-rich	Proline, serine and threonine-rich
Rab	Ras-related protein
RPN2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2
RUSH	Retention using selective hooks
SARS-CoV	Severe acute respiratory syndrome coronavirus
Scyl1	N-terminal kinase-like protein
SDS	Sodium dodecyl sulfate
sgRNA	Guide RNA
SI	Sucrase-isomaltase
SLC27A	Long-chain fatty acid transport protein
SNARE	SNAP receptor

SOB	Super optimal broth
ST3GAL4	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4
ST6GAL1	Beta-galactoside alpha-2,6-sialyltransferase 1
STL	Solanum tuberosum lectin
TGN	Trans-Golgi network
TM9SF	Transmembrane 9 superfamily
TMD	Transmembrane domain
TMED10	p24 family protein delta-1
TMED2	p24 family protein beta-1
TMEM55	Phosphatidylinositol 4,5-bisphosphate 4-phosphatase
TPST1	Protein-tyrosine sulfotransferase 1
TPST2	Protein-tyrosine sulfotransferase 2
Ugt35a	UDP-glycosyltransferase family 35 member A1
UXS1	UDP-glucuronic acid decarboxylase 1
VVA	Vicia villosa lectin
WFA	Wisteria floribunda lectin
ZFPL1	Zinc finger protein-like 1

1. Introduction

1.1 The Golgi apparatus

The Golgi apparatus is an organelle which sits at the heart of the endomembrane system of eukaryotic life. In most mammalian cells, the Golgi resembles several flattened cisternae which are stacked adjacent to one another and are situated in the vicinity of the nucleus. The Golgi is a heterogenous polarised structure whose different cisternae and domains can be broadly categorised according to their position in space and their protein and lipid constituents. The cisterna facing the endoplasmic reticulum (ER) exit sites (ERES) is termed the cis-Golgi (Gomez-Navarro and Miller, 2016). The cis-Golgi receives membrane and protein cargo in the form of COPII-coated vesicles which bud from the ERES (Gomez-Navarro and Miller, 2016). These vesicles are transported along microtubules and are eventually tethered to the cis-Golgi target membrane prior to SNARE-mediated membrane fusion (Gomez-Navarro and Miller, 2016). The cis-Golgi can also act as a donor membrane in which retrograde COPI vesicles serve to return membranes and ER-resident escapees to the ER (Gomez-Navarro and Miller, 2016), a process which will be discussed in detail later. At the interface of the ERES and the cis-Golgi is the ER-Golgi intermediate compartment (ERGIC), a dynamic compartment which is involved in the bilateral exchange of membranes and cargo between the neighbouring organelles (Appenzeller-Herzog and Hauri, 2006). Neighbouring the cis-Golgi, distal from the ERES, are the cisternae which make up the medial-Golgi and the trans-Golgi. Adjacent to the trans-Golgi is the trans-Golgi network (TGN) which resembles a dynamic complex of tubules and vesicles (Pantazopoulou and Glick, 2019). The TGN is involved in the sorting and exchange of proteins and membranes with post-Golgi compartments such as the plasma membrane and the endolysosomal system (Pantazopoulou and Glick, 2019).

The Golgi functions as a central sorting hub for membrane traffic within the cell. It is involved in the exchange of a wide array of efferent and afferent membrane carriers between the many compartments of the endomembrane system. Another critical function of the Golgi is in the enzymatic modification of lipids and proteins. One such modification of great significance is glycosylation. The interplay between two of the core functions of the Golgi, membrane trafficking and glycosylation, will underpin the focus of this thesis.

1.2 Glycan-modifying enzymes

Secreted and membrane proteins are synthesised at the ER where they are integrated into the ER membrane or released into the ER lumen (Higy et al., 2004). In the ER, newly synthesised proteins undergo protein folding and the addition of the N-linked glycan core structure (Moremen et al., 2012). Glycosylation involves the modification of specific protein residues with a wide range of different carbohydrate moieties, often acting in pathways of sequential modification (Moremen et al., 2012). Glycosylation modifications can exhibit a vast degree of heterogeneity and can have significant implications for a protein's stability or function (Rudd and Dwek, 1997). As proteins progress through the Golgi apparatus, they are subjected to processing of N-linked glycans and the addition of O-linked glycans (Moremen et al., 2012). This is mediated by a large number of different glycosylation enzymes which display variable distributions across the Golgi cisternae.

Golgi-resident glycosylation enzymes belong to two subcategories: glycosyltransferases and glycosidases (Lombard et al., 2014). These enzymes exhibit substantial diversity, not only across eukaryota but within eukaryotic genomes (Wang et al., 2017). For example, humans have over 300 genes encoding glycosylation enzymes while a plant genome can contain well in excess of 400 such genes (Lombard et al., 2014; Nikolovski et al., 2012; Wang et al., 2017). A near universal feature of Golgi enzymes is that they are type II transmembrane proteins, characterised by having a short cytoplasmic tail at their N terminus (generally 5– 20 amino acids), a single transmembrane domain (TMD) and a disordered stem/stalk region which acts as a spacer between the bilayer and the luminally facing catalytic domain. Many Golgi enzymes are able to form homo or heterodimers through interactions via their TMDs and lumenal domains (Hartmann-Fatu et al., 2015; McCormick et al., 2000; Nilsson et al., 1996). The catalytic domains have diverse sequences and structures and are responsible for enzymatic and substrate specificity (Wang et al., 2017). The region comprising the cytoplasmic tail, TMD and stem/stalk has been termed the CTS domain and it is this domain that is responsible for the Golgi-targeting of these enzymes (Figure 1.1; Tu and Banfield, 2010). Over decades of research, each component of the CTS domain has been implicated in distinct Golgi-targeting mechanisms. These different CTS domains are also responsible for the specific sub-Golgi distributions of different enzymes. It is worth noting that particular enzymes are generally not restricted to single cisternae but are spread over 2–3 cisternae with their positions broadly corresponding to their order in their pathway of action. While the CTS domain is clearly sufficient for Golgi-localisation, what is uncertain is the relative contribution of the different parts of the CTS domain.

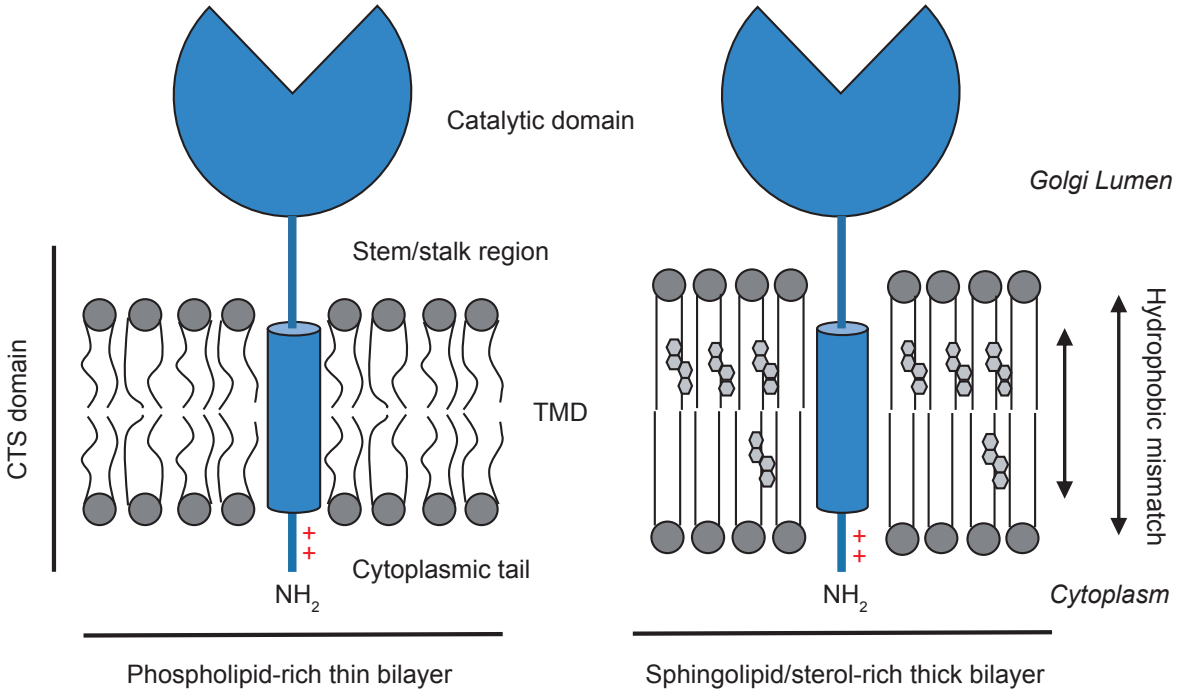


Figure 1.1 Topology of Golgi enzymes and their residence in membranes of varying lipid content and thickness. The 'CTS' domain corresponds to the cytoplasmic tail, TMD and stem/stalk region. The residues of the N-terminal cytoplasmic tail are characteristically basic and can interact with coatomer directly or indirectly. The TMDs of Golgi enzymes are relatively short and consequently favour residence in membranes that are phospholipid-rich, disordered and hence thin. In contrast, the acyl chains of sphingolipid/sterol-rich membranes are saturated resulting in greater order, tighter chain packing and therefore greater bilayer thickness. Residence in thicker, sphingolipid/sterol-rich, membranes is energetically unfavourable due to an incompatibility between the width of the bilayer and the length of the hydrophobic TMD.

Any model for how Golgi enzymes are retained in the Golgi has to be integrated into what is known about membrane traffic within the Golgi. Transport vesicles leave the TGN for various destinations and so Golgi enzymes will need to avoid entering these along with cargo. In addition, there are transport vesicles that move between cisternae, and it is now widely believed that these carry Golgi enzymes back to earlier cisternae as cargo is moved forward in cisternae that progress toward the TGN – the cisternal maturation model (Figure 1.2). This model posits that the Golgi cisternae are constantly maturing or progressing from a cis-Golgi cisterna to a trans-Golgi cisterna (Glick and Nakano, 2009). As cisternae mature, they develop a more 'trans-like' profile in terms of their protein and lipid content (Glick and Nakano, 2009). In order to maintain the polarity in cisternal content across the Golgi, Golgi-residents are selectively recycled back to earlier cisternae in vesicles generated by COPI, a vesicle coat made up of coatomer subunits. Cisternal maturation in combination with a vesicular counter-flow serves to provide stasis and ensure the correct localisation of Golgi enzymes (Glick and Nakano, 2009). Thus, the Golgi enzyme sorting machinery must not just ensure that Golgi enzymes are retained within the organelle, but also that the different enzymes are distributed to a differing subset of cisternae. The remainder of this chapter will discuss the role that the cytoplasmic tails and TMDs of Golgi-resident glycan-modifying enzymes play in their sorting in the context of the cisternal maturation model.

1.3 Transmembrane domain-dependent sorting

1.3.1 Membrane thickness model

Initially the TMDs of a handful of Golgi enzymes were shown to be important for their Golgi-localisation, despite there being no discernible consensus in their primary sequences (Bretscher and Munro, 1993; Munro, 1991; Munro, 1995; Nilsson et al., 1991; Swift and Machamer, 1991). An initial comparison, based on a limited collection of available sequences, showed that Golgi enzymes exhibited shorter TMDs when compared to plasma membrane-resident type II transmembrane proteins (Bretscher and Munro, 1993). Subsequent analysis, conducted on an expansive dataset of fungal and vertebrate sequences, showed that across eukaryotes, TMD length of bitopic membrane proteins generally increased in correlation with later residence in the secretory pathway (Sharpe et al., 2010). The most striking difference was observed when comparing the length of the TMDs of ER/Golgi-residents versus post-Golgi-residents. Furthermore, it was shown that there is a preference for bulkier residues in the exoplasmic half of the TMDs of Golgi-residents when compared to that of their plasma membrane counterparts (Sharpe et al., 2010). It is not yet fully resolved whether the plasma membrane really is thicker than the

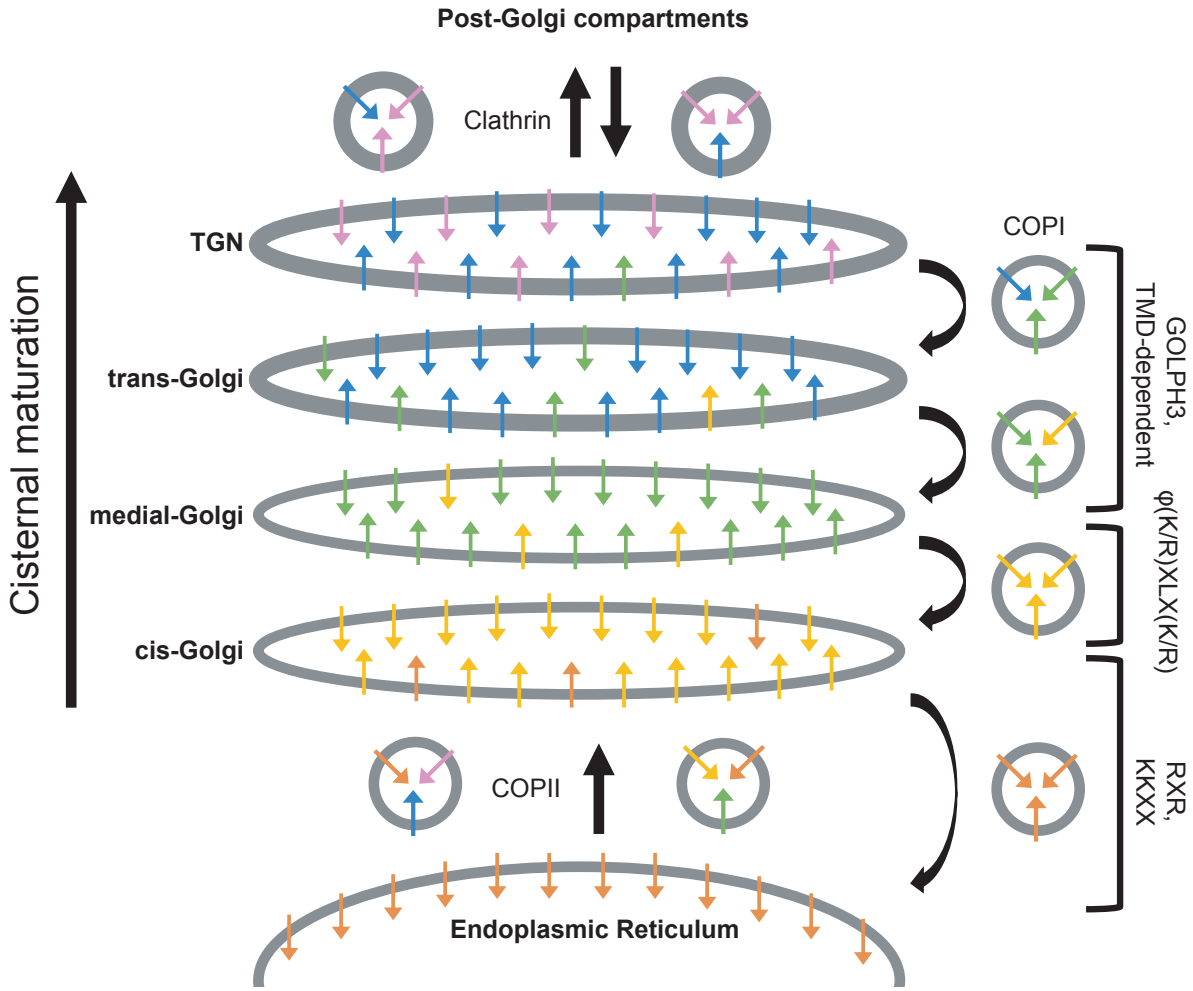


Figure 1.2 Cisternal maturation model. Golgi cisternae progressively mature from cis to trans as they gradually transition from phospholipid-rich membranes containing cis-Golgi resident proteins to sphingolipid/sterol-rich membranes containing trans-Golgi residents. In order to maintain this polarity, lipids and Golgi-residents are subject to selective retrograde trafficking in COPI vesicles to their correct, earlier, cisternae of residence. This selective vesicular transport is likely to be mediated by several, nonmutually exclusive, mechanisms. It is likely that these different mechanisms are complementary and are of different prominence at different places within the Golgi stack. ER-resident cargo proteins are targeted for COPI-dependent retrograde transport from the Golgi to the ER by cytoplasmic KKXX or RXR ER-retrieval sequences. For intra-Golgi vesicular traffic, several cis-Golgi enzymes harbour a φ(K/R)XLX(K/R) consensus motif through which they interact directly with the δ/ζ1-COP subunits of coatomer. It is likely that the φ(K/R)XLX(K/R) motif is important for recycling within the early Golgi. At the late Golgi, GOLPH3 is recruited to the membrane by PtdIns4P where it can simultaneously interact with coatomer and the cytoplasmic tails of a different subset of Golgi enzymes to promote their segregation into COPI vesicles. Also at the late-Golgi, an increase in sphingolipid/sterol content results in a concomitant increase in membrane thickness. This favours the segregation of Golgi enzymes with short TMDs into phospholipid-rich budding COPI vesicles. In contrast, at the TGN clathrin-coated vesicles destined for post-Golgi compartments are enriched in sphingolipids/sterols which serves to exclude Golgi enzymes with short TMDs from these carriers.

membranes of the ER and Golgi, but such an increase has at least been seen by electron microscopy (EM) including a study based on EM of frozen unfixed cells (Bykov et al., 2017; Mitra et al., 2004; Schneiter et al., 1999). It was thus hypothesised that the short TMDs of the Golgi enzymes were retained at the Golgi as a result of their favoured residence in thinner membranes of the early secretory pathway (Bretscher and Munro, 1993). In other words, it is more energetically favourable to have a short hydrophobic α-helix residing in a compatibly thin hydrophobic bilayer such as that of the ER and Golgi than in a thicker post-Golgi bilayer (Figure 1.1). It is worth noting that this hydrophobic mismatch of a long TMD in a short bilayer is thought to be more tolerable and energetically favourable than that of a short TMD in a thicker bilayer (Killian, 1998; Phillips et al., 2009). This could account for the traffic of biosynthetic cargo with long TMDs which are synthesised and inserted into the membrane at the ER and must traverse the secretory pathway to post-Golgi destinations.

The short TMDs of Golgi enzymes may favour residence in a thinner membrane of the early secretory pathway, but what would cause a difference in membrane thickness? The lipid compositions of different membranes are not uniform throughout the secretory pathway (van Meer, 1989). Membranes of the ER are characteristically phospholipid-rich with low levels of sphingolipids and sterols (van Meer, 1998). However, while cis-Golgi cisternae are broadly comparable to ER membranes in their lipid composition, there is an increase in sphingolipid and sterol content in the trans-Golgi and on to post-Golgi compartments (Holthuis et al., 2001). This variation in membrane composition is driven in part by a number of factors. While cholesterol is synthesised at the ER or delivered to the ER from lysosomes, most is then transported to later secretory compartments via non-vesicular transport (Baumann et al., 2005; Graham and Burd, 2011; Prinz, 2007). Furthermore, the localisation of sphingolipid synthesis enzymes at the trans-Golgi and plasma membrane can account for a raised sphingolipid content at these sites (van Meer et al., 2008). Moreover, the gradient in phospholipids, sphingolipids and sterols could be driven by the selective vesicular transport of lipids between Golgi cisternae in the context of cisternal maturation (Figure 1.2; Holthuis et al., 2001). In comparison to phospholipid-rich membranes, sphingolipid and sterol-rich membranes exhibit greater acyl chain order and bilayer thickness (van Meer et al., 2008). Therefore, with the change in lipid content throughout the secretory pathway there should be a concomitant increase in bilayer thickness and order, and it has also been proposed that longer TMDs could themselves contribute to thickening the bilayer (Bykov et al., 2017; Phillips et al., 2009; Sharpe et al., 2010). It is this variation in bilayer thickness throughout the Golgi which lends itself to the retention of its resident enzymes through their relatively short TMDs (Bretscher and Munro, 1993).

The cisternal maturation model states that as cisternae progress from cis to trans, Golgi-residents are cycled back via COPI-dependent vesicular transport (Glick and Nakano, 2009). How would the membrane thickness model work in the context of cisternal maturation? Experiments with liposomes show that at the appropriate concentrations, sphingolipids and sterols segregate away from phospholipid-rich regions by phase separation (Dietrich et al., 2001; Roux et al., 2005). If this occurred in vivo, it could result in the formation of a lipid domain which favours residence of Golgi enzymes and another which favours cargo proteins destined for post-Golgi compartments (Patterson et al., 2008). It is worth noting that this lipid segregation is unlikely to occur at the ER and cis-Golgi owing to the low levels of sphingolipids and sterols (Holthuis et al., 2001). However, it is also possible that lipid segregation is driven by high curvature such as that found at Golgi rims and during the biogenesis of COPI vesicles (Brugger et al., 2000; Patterson et al., 2008; Roux et al., 2005). Ordered lipid domains consisting of sphingolipids and sterols favour flat cisternal membranes and are less tolerant of curvature than the relatively disordered phospholipid-rich lipid domains (Campelo et al., 2017). This phenomenon has been demonstrated in giant unilamellar vesicles composed of sphingolipids, sterols and phospholipids through the application of high curvature via the extrusion of nanotubules (Roux et al., 2005). It has been reported that COPI vesicles formed in vitro from isolated mammalian Golgi stacks have lower levels of sterols and sphingolipids, and more phospholipids, than the Golgi from which they budded (Brugger et al., 2000). This would be consistent with the vesicles preferentially selecting ER-like lipids, although a caveat is that the lipid composition of the Golgi was determined for the whole stack whereas the vesicles may have been budding from only a subset of cisternae. Such selection would certainly favour the Golgi enzymes, with their short TMDs, segregating into the budding COPI vesicle (Holthuis et al., 2001). It is worth noting that unlike COPI, clathrin-coated vesicles at the TGN are actually enriched for sphingolipids and sterols which therefore may serve to exclude Golgi enzymes from post-Golgi carriers (Klemm et al., 2009). Further support for the importance of lipids in the retention of Golgi-residents comes from the findings that genetic or chemical disruption of sphingolipid synthesis causes the mislocalisation of Golgi enzymes (van Galen et al., 2014; Wood et al., 2012).

A difference between the membrane thickness of COPI vesicles and that of their cisternal membrane of origin has long been observed in fixed samples (Orci et al., 1996). Furthermore, new evidence has arisen from applying EM tomography to cryopreserved *Chlamydomonas reinhardtii* (Bykov et al., 2017). The tomograms reveal a clear difference between the membrane thickness of budding COPI vesicles and cisternal membranes at the medial/trans-Golgi (Bykov et al., 2017). Interestingly, while there appears to be no

difference in membrane thickness between the cis and medial-Golgi, there is a notable increase in thickness between the medial and trans-Golgi (Bykov et al., 2017). This would suggest that TMD-dependent sorting at the Golgi primarily occurs at the medial/trans interface with there being insufficient levels of sphingolipids at the cis-Golgi to affect the bilayer or promote lipid segregation (Figure 1.2; Holthuis et al., 2001). In contrast, at the trans-Golgi the sphingolipid content is sufficient for segregation to occur due to the local presence of sphingolipid synthases (Figure 1.3A; van Meer et al., 2008). This would suggest that the recycling of enzymes from early-Golgi cisternae occurs through mechanisms independent of the physiochemical properties of the TMD.

1.3.2 Sequence-specific transmembrane domain sorting

Studies on the role of the TMDs in Golgi retention have examined only a limited range of proteins but most have suggested that the key aspect for retention is the physiochemical properties of the TMD. However, recent studies conducted in plants highlighted a glutamine residue in the TMD of N-acetylglucosaminyltransferase I (GnTI) which was important for its targeting to the cis/medial-Golgi (Schoberer et al., 2019). Mutagenesis of this residue caused GnTI to localise later in the Golgi and be directed to the vacuole for degradation (Schoberer et al., 2019). It remains to be seen whether this residue regulates the localisation of the enzyme through an interaction with the membrane-spanning domain(s) of a proteinaceous COPI adaptor. On that note, a recent flurry of CRISPR-Cas9 screens in mammalian cells with readouts linked to glycosylation status has implicated the integral membrane protein TM9SF2 in the retention of a subset of Golgi enzymes (Pacheco et al., 2018; Tanaka et al., 2017; Tian et al., 2018; Yamaji et al., 2019). TM9SF2 is a Golgi-localised nine transmembrane-spanning protein of unknown function which belongs in a family with three other proteins, all of which have a variety of orthologs across eukaryota (Woo et al., 2015). Notably, all four TM9SF proteins have been detected in COPI vesicles generated from semi-intact cells (Adolf et al., 2019). Knockout of TM9SF2 causes a global defect in glycosylation and the mislocalisation of various Golgi enzymes (Pacheco et al., 2018; Tanaka et al., 2017; Tian et al., 2018; Yamaji et al., 2019). Furthermore, TM9SF2 was reported to interact with the enzyme A4GALT in vitro (Yamaji et al., 2019). Interestingly, it was shown that TM9SF2 has a conserved KxD/E COPI-binding motif at its C-terminus and disruption with point mutations or a C-terminal tag was sufficient to cause the protein to accumulate in post-Golgi compartments (Woo et al., 2015). Could TM9SF2 and its relatives act as novel COPI adaptors by simultaneously interacting with the COPI coat and sequence motifs in the TMDs of Golgi enzymes? This would be analogous to the COPI adaptor Rer1 which detects ER-retrieval signals in the TMDs of a number of ER-resident proteins (Sato et al., 2001).

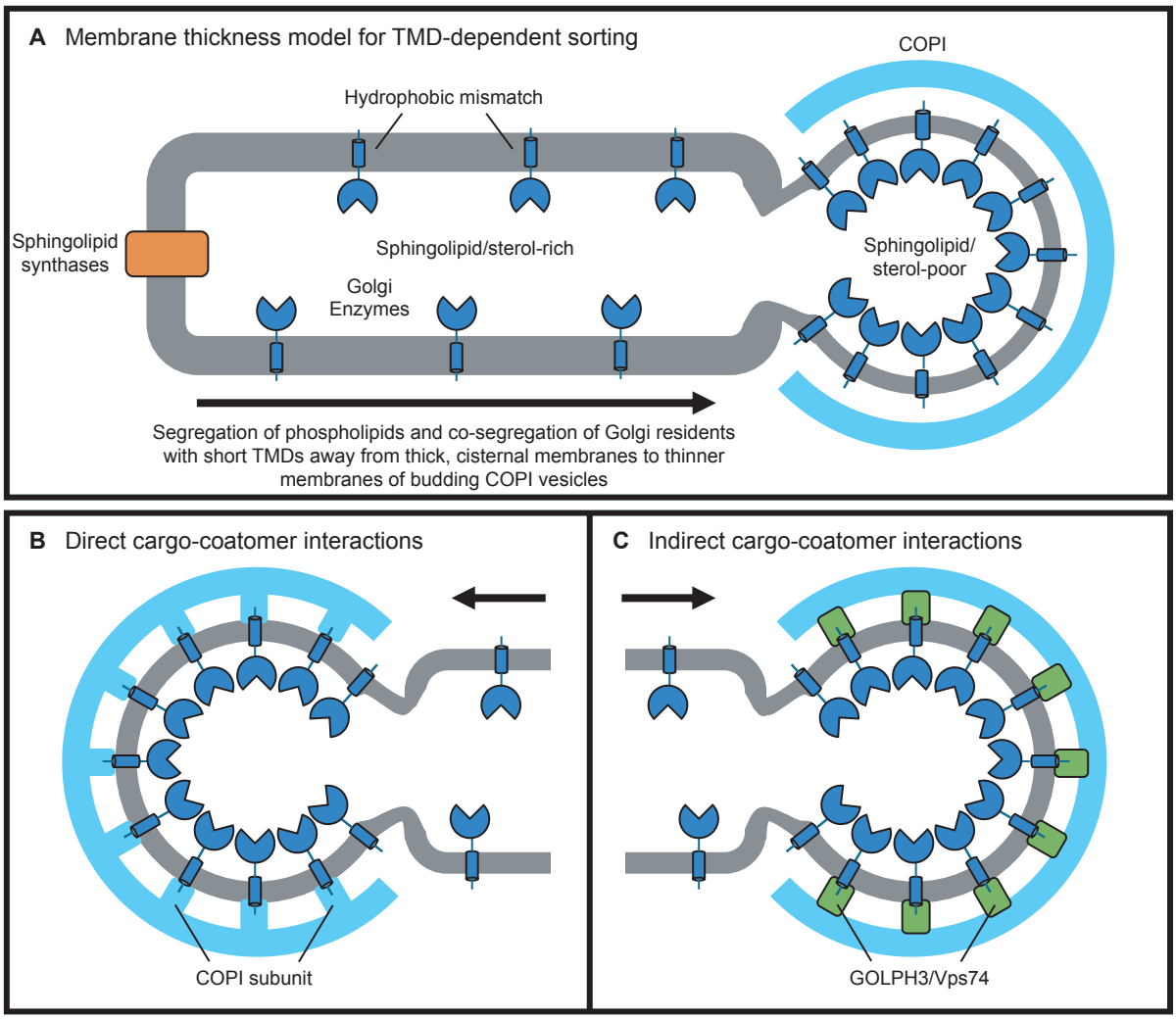


Figure 1.3 Possible mechanisms of Golgi enzyme retention. (**A**) Sphingolipid synthases and selective vesicular transport of lipids drive a local increase in sphingolipid levels at the late-Golgi. The high curvature, or enrichment of short TMD proteins, associated with budding COPI vesicles may promote the segregation of phospholipids away from the sphingolipids and sterols of the cisterna and into the vesicular bud. It is subsequently energetically favourable for Golgi enzymes, with their relatively short TMDs, to segregate into the thinner phospholipid-rich membrane of the vesicular bud. Segregation of cargo can also be driven by the direct interaction of the cytoplasmic tails of the enzymes with various subunits of the COPI coat (**B**), or occur indirectly via COPI adaptors such as GOLPH3/Vps74 (**C**)

1.4 Cytoplasmic tail sorting motifs

1.4.1 Vps74/GOLPH3

Some of the early studies on Golgi enzyme retention that found a role for the TMD also found that the cytoplasmic tail could contribute to retention (Burke et al., 1992; Burke et al., 1994; Munro, 1991; Nilsson et al., 1991). In some cases, the cytoplasmic tail on its own could impart a Golgi-sorting signal but the mechanism by which this signal was recognised was unclear (Burke et al., 1994; Osman et al., 1996). This changed with the discovery that the yeast protein Vps74, a peripheral Golgi protein, could simultaneously interact with the cytoplasmic tails of a number of yeast mannosyltransferases and with coatomer (Figure 1.3C; Tu et al., 2008; Schmitz et al., 2008). In the absence of *Vps74*, some Golgi-resident enzymes were mislocalised to the vacuole and glycosylation was perturbed (Schmitz et al., 2008; Tu et al., 2008). *Vps74* is conserved between fungi and metazoans and there are two mammalian orthologs: *GOLPH3* and *GOLPH3L*. The mammalian orthologs can partially rescue the phenotypes of yeast lacking *Vps74* (Tu et al., 2008). GOLPH3/Vps74 proteins are reported to interact with the δ-COP and β-COP components of the COPI coat through a conserved cluster of arginine residues within a short (~50 amino acids) unstructured region near their N-termini (Tu et al., 2008; Tu et al., 2012). Furthermore, the mammalian and yeast proteins share the ability to interact with the phosphoinositide lipid PtdIns4P which appears to be essential for their Golgi-localisation (Dippold et al., 2009; Wood et al., 2009; Wood et al., 2012). Levels of PtdIns4P increase from the cis- to the trans-Golgi (Cheong et al., 2010). It was demonstrated that GOLPH3/Vps74 proteins are recruited to the late-Golgi by PtdIns4P where they can interact with the cytoplasmic tails of Golgi enzymes. It is here that GOLPH3/Vps74 facilitates the retrograde transport of enzymes from late cisterna to the earlier Golgi compartments in which they normally reside (Wood et al., 2012). In other words, GOLPH3 can be likened to a night club bouncer at the entrance to the post-Golgi compartments where it selectively sends Golgi enzymes backwards against the flow of the maturing queue of cisternae.

The consensus motif recognised by Vps74 was found to be (F/L)(L/I/V)XX(R/K) (Tu et al., 2008); however, the recognition motif is less well-defined for GOLPH3 and GOLPH3L. Until recently, only a handful of enzymes have been shown to interact with GOLPH3 or have been shown to require GOLPH3 for their proper Golgi localisation (Table 1.4.1) (Ali et al., 2012; Eckert et al., 2014; Isaji et al., 2014; Pereira et al., 2014). A non-peer-reviewed report was recently made available in a preprint repository which characterises novel GOLPH3 clients and suggests a GOLPH3 consensus motif (Rizzo et al., 2019). Since this a

preprint report and was made available after the initial publication of Table 1.4.1 in a review (Welch and Munro, 2019), it will therefore be discussed later (Chapter 4).

Table 1.4.1 Reported direct and indirect interactions between the cytoplasmic tails of Golgi enzymes and coatomer

Golgi Enzyme	Cytoplasmic tail sequence	Reported interactor(s)
EXT1*	MQAKKRY	GOLPH3 (Chang et al., 2013)
EXT2*	MCASVKYNIRGPALIPRMKTKHRIY	GOLPH3 (Chang et al., 2013)
GALNT3	MAH**LKRLVK**LHIKRHYHKK	δ-COP, ζ1-COP (Liu et al., 2018)
GALNT4	MAVRWTWAGKSC	δ-COP (Liu et al., 2018)
GALNT6	**MRLLRR**RH	δ-COP (Liu et al., 2018)
GALNT8	MMF**WRKLPK**	δ-COP, ζ1-COP? (Liu et al., 2018)
GALNT12	MWGRTARRRCPRELRRGRE	GOLPH3 (Pereira et al., 2014)
GCNT1	M**LRTLL**RRR	GOLPH3 (Ali et al., 2012; Eckert et al., 2014); δ-COP, ζ1-COP (Liu et al., 2018)
GNPTAB	M**LFKLLQR**QTYTCLSHRYGLY	δ-COP, ζ1-COP Liu et al., 2018)
POMGNT1	MDDWKPSPLIKPFGARKKRSWYLTWKYKLTNQRALRR	GOLPH3 (Pereira et al., 2014)
ST6GAL1	MIHTNLKKK	GOLPH3 (Eckert et al., 2014; Isaji et al., 2014)
ST3GAL4	MVSKSRWK	GOLPH3 (Isaji et al., 2014)

All cytoplasmic tail sequences represent the extreme N-terminus preceding the TMD as assigned according to UniProt with the exception of GALNT8 (Bateman et al., 2019). Note that annotations of TMDs are not always precise so ends of tail sequences should be viewed with caution. Underlined residues highlight a proposed GOLPH3 recognition motif characterised by a membrane-proximal consensus sequence consisting of a hydrophobic residue upstream of a di-basic motif (Maccioni et al., 2011). Residues in bold highlight a φ(K/R)XLX(K/R) motif corresponding to a δ-COP/ζ1-COP consensus binding motif (Liu et al., 2018). Highlighted residues correspond to a $Wx_{n(1-6)}$[W/F] motif predicted to bind to δ-COP (Suckling et al., 2015). *Note that EXT1 and EXT2 form heterodimers so the ability of either tail to interact directly with GOLPH3 is not clear (McCormick et al., 2000).

The work of several studies came to the consensus that their cytoplasmic tails share a motif loosely characterised as a membrane-proximal di-basic stretch preceded by a hydrophobic residue (Ali et al., 2012; Eckert et al., 2014; Isaji et al., 2014; Pereira et al., 2014). For example, GCNT1 was shown to interact with GOLPH3 via an LLLRR motif in its cytoplasmic tail, as was ST6GAL1 which has a similar LKK motif (Ali et al., 2012). GOLPH3 is required for the incorporation of GCNT1 and ST6GAL1 into COPI vesicles in vitro, a requirement which was not shared with other COPI cargo such as B4GALT1 (Eckert et al., 2014). In *Drosophila melanogaster*, the orthologous GOLPH3 protein was found to interact

with, and be important for the trafficking of, the exostosin enzymes EXT1, EXT2 and EXTL3 (Chang et al., 2013). It was also shown that the interaction of GOLPH3 with EXT1 and EXT2 was conserved in mammalian cells (Chang et al., 2013). It is noteworthy that EXT1 and EXT2 form heterodimers (McCormick et al., 2000). In fact, several Golgi enzymes have been shown to heterodimerise with each other and possibly even form higher order oligomers (Hartmann-Fatu et al., 2015; Jungmann and Munro, 1998; Kellokumpu et al., 2016). Therefore, caution is needed when trying to identify sorting motifs in cytoplasmic tails given the possibility of enzymes interacting indirectly with GOLPH3 as part of a heterodimer. When considering the increase in complexity in mammalian Golgi enzyme biology relative to that of yeast it seems likely that only a small proportion of GOLPH3 clients are currently known. Interestingly, a bioinformatic analysis of Golgi-resident type II transmembrane proteins searching for a membrane-proximal di-hydrophobic, di-basic motif suggested the motif was present in 15% of proteins tested (33 enzymes) (Maccioni et al., 2011). Further experimental interrogation is required to inform and compliment future bioinformatic analyses in order to elucidate the GOLPH3 recognition motif and complete client list.

1.4.2 Direct interactions between enzymes and coatomer

Several direct interactions have been described between the COPI coat and its cargo, many of which are associated with retrieval of ER residents from the cis-Golgi for return to the ER, as opposed to intra-Golgi trafficking of Golgi enzymes (Figure 1.2). One of the best characterised ER-retrieval motifs is the dilysine motif: a membrane proximal KKXX or KXKXX sequence (Ma and Goldberg, 2013; Nilsson et al., 1989). The motif is found at the cytoplasmic C-terminus of many ER-resident type I transmembrane proteins and interacts directly with the α- and β'-COP subunits (Ma and Goldberg, 2013). Another ER-retrieval motif is the RXR motif which is found in the cytoplasmic tails of integral membrane proteins of varying topology including multi-pass proteins that are components of multimeric complexes (Michelsen et al., 2005). Often, complete assembly of the multimeric complex masks the RXR motif and so suppresses ER-retrieval (Michelsen et al., 2005). This serves to limit the premature progression of improperly assembled complexes through the secretory pathway. RXR motifs can also be masked by post-translational modification or through the binding of regulatory proteins (Michelsen et al., 2005). The position of the RXR motif in relation to the membrane appears critical for its function (Shikano and Li, 2003). This can be illustrated for type II transmembrane proteins by comparing two different glycosylation enzymes. The poly-arginine stretch ^{3}RGERRRR^{9} in the cytoplasmic tail of human MOGS is positioned 29 residues distal to the membrane and is sufficient to target

the protein to the ER (Boulaflous et al., 2009). In contrast, a similar ^{1}MRRRSR^{6} motif positioned immediately proximal to the membrane in the tail of GALNT2 is sufficient for Golgi-targeting (Liu et al., 2018). The molecular identity of the receptors and/or coat components involved in either arginine-based interaction is unclear but it would be surprising if it was not COPI-dependent. This intriguing bifurcation of the sorting signal, apparently based on membrane proximity alone, suggests the positioning of the cytoplasmic tails in the context of the structure of coatomer is functionally critical.

Recently a direct interaction between the cytoplasmic tails of Golgi enzymes and coatomer has been reported (Figure 1.3B; Liu et al., 2018). Based on a previous observation that disease mutations in the cytoplasmic tail of GNPTAB caused its mislocalisation from the Golgi to the lysosome, Liu and colleagues used a proximity-dependent labelling approach to find cytoplasmic tail-interacting proteins (Liu et al., 2018; van Meel et al., 2014). They leveraged the topology of GNPTAB, a double-pass enzyme with a cytoplasmic N- and C terminus, by placing the promiscuous biotin ligase BioID2 on the C-terminus so as to biotinylate proteins that interacted with the N terminus (Liu et al., 2018). After isolating and sequencing the interacting proteins, they found that, amongst others, the δ-COP and ζ1-COP subunits of coatomer were top hits when comparing the wild-type protein to the disease mutants (Liu et al., 2018). Notably GOLPH3 was absent from the list of hits suggesting that the interaction with coatomer was GOLPH3-independent (Liu et al., 2018). The cytoplasmic tail of GNPTAB and various cis-Golgi resident enzymes were found to bind directly to the μ-homology domain (MHD) of δ-COP while some also bound to ζ1-COP (Table 1.4.1; Liu et al., 2018). Interestingly, the tails interacted with the binding pocket on δ-COP MHD that is structurally analogous to the binding pocket of the μ subunit of the AP-2 clathrin adaptor which recognises the endocytic sorting motif Yxxφ in cytoplasmic tails (Liu et al., 2018; Owen and Evans, 1998). The δ-COP/ζ1-COP consensus binding motif was identified as φ(K/R)XLX(K/R) and it was shown that mutagenesis of this motif was, in most instances, sufficient to both ablate coatomer binding in vitro and disrupt Golgi-localisation in vivo (Liu et al., 2018). Curiously, GCNT1, which was previously reported to interact with coatomer indirectly via GOLPH3, was also found to interact directly with coatomer (Ali et al., 2012; Eckert et al., 2014; Liu et al., 2018). The cytoplasmic tail of GCNT1, appears to have a putative GOLPH3 binding site that overlaps the δ-COP/ζ1-COP binding motifs, suggesting redundancy in the sorting signals in the cytoplasmic tail of at least one Golgi enzyme (Table 1.4.1), and disruption of both is required to ablate an interaction with coatomer (Liu et al., 2018). However, it should be noted that in this study deletion of *GOLPH3* did not affect the Golgi localisation of GCNT1, in contrast to what had been reported previously (Ali et al., 2012; Liu et al., 2018). However, neither study examined

GOLPH3's closely related paralog *GOLPH3L*. The binding properties of this protein have not been reported, but if it does bind GCNT1 it may be that its expression level varies between cell lines.

Liu and colleagues also demonstrated in vitro binding between a peptide corresponding to the cytoplasmic tail of GALNT4 and δ-COP MHD via a $Wx_{n(1-6)}$[W/F] motif to a binding pocket distinct from that which recognises φ(K/R)XLX(K/R) in other Golgi enzymes (Liu et al., 2018; Suckling et al., 2015). The $Wx_{n(1-6)}$[W/F] motif is an evolutionarily conserved δ-COP MHD-interacting motif which has been described in a number of proteins involved in COPI biology including ARFGAP1, Scyl3 (Cex1 in yeast) and Dsl1 (Andag and Schmitt, 2003; Cosson et al., 1998; Rawet et al., 2010; Suckling et al., 2015). It remains to be seen whether this interaction occurs in vivo with the full coat complex and the tail in the context of a membrane. Various structures of coatomer suggest there may be sufficient flexibility in δ-COP and the COPI coat as a whole to accommodate for a direct interaction with the cytoplasmic tails (Bykov et al., 2017; Dodonova et al., 2015; Dodonova et al., 2017). Another curiosity is that $Wx_{n(1-6)}$[W/F] motifs often feature amongst highly acidic stretches (Suckling et al., 2015); however, this does not appear to be the case for the motif in the tail of GALNT4. Regardless, this potentially presents a novel COPI cargo-sorting motif in the context of Golgi enzymes. For example, the motif appears to be present in the tails of POMGNT1 (Table 1.4.1) and a number of human fucosyltransferases, while a disease-causing missense mutation in B3GALT6 appears to generate a synthetic $Wx_{n(1-6)}$[W/F] motif in its cytoplasmic tail (Nakajima et al., 2013).

All the enzymes that have been shown to bind directly to coatomer are cis-Golgi residents while several of the GOLPH3 clients are localised later in the mammalian Golgi stack. Are these different sorting signals important for segregating different enzymes into vesicles that bud from different cisternae (Figure 1.3)? In other words, are enzymes with a δ-COP/ζ1-COP-binding motif sorted from earlier cisterna, such as the medial-Golgi, while those with a GOLPH3-binding motif are sorted from the late-Golgi? This may result in distinct subclasses of COPI vesicles, defined by their cisterna of origin and their variable cargo loads, which serve to fine-tune the localisation of enzymes across the Golgi cisternae. Enzymes like GCNT1 will be retained by both sorting motifs and benefit from redundant mechanisms to keep them from the late-Golgi. What is not clear is how the distinction is achieved between COPI vesicles associated with intra-Golgi traffic and those associated with ER-retrieval. These different subtypes of vesicle must differ in cargo content yet both have been proposed to select cargo by direct interactions between cargo and the COPI coat subunits. It is conceivable that there are differences in the conformation of the COPI

coat in different sub-Golgi sites to facilitate distinct modes of direct cargo interaction and selection.

1.5 The membrane thickness model and tail-dependent sorting

So how do cytoplasmic tail-dependent mechanisms of Golgi-enzyme sorting fit alongside mechanisms of TMD-based retention based on the membrane thickness model? It seems quite possible that Golgi enzymes have multiple distinct, complimentary and additive mechanisms for Golgi-targeting (Figure 1.3). The short TMDs of the Golgi enzymes could promote their segregation into the thinner and less ordered membranes of budding COPI vesicles. This would occur at the late-Golgi where the sphingolipid and cholesterol content is sufficiently high that most of the bilayer is thicker and ordered. The direct and/or indirect interaction of the cytoplasmic tails with coatomer then serves to contribute to this segregation, either by promoting segregation in its own right or by stabilising those which have segregated as a result of lipid composition. Once budding is complete, the vesicle loaded with Golgi enzymes and phospholipids can fuse with an earlier cisterna in order to maintain the correct lipid and protein distribution between the Golgi cisternae. We suggest that TMD-dependent and GOLPH3-dependent sorting occurs predominately at the late-Golgi while direct tail interactions with the δ/ζ1-COP subunits are important for sorting within the earlier part of the Golgi. At the cis-Golgi; dilysine, diarginine and other ER-retrieval mechanisms are important for retrograde trafficking from the Golgi to the ER. Evidence for a complimentary relationship between tail-dependent and TMD-dependent sorting is that genetic disruption of *GOLPH3*/*Vps74* and genetic or chemical disruption of sphingolipid biosynthesis result in similar defects in Golgi enzyme sorting (van Galen et al., 2014; Wood et al., 2012). Evidence for an additive relationship between the two mechanisms is that *Vps74* has negative genetic interactions with several genes involved in sphingolipid and phospholipid synthesis (Wood et al., 2012).

1.6 Turnover of mislocalised Golgi enzymes

A common feature across eukaryotic systems is that the loss of retention of Golgi enzymes, whether through perturbations in cytoplasmic tail- or TMD-dependent targeting mechanisms, often results in their targeting to the lysosome for degradation (Chang et al., 2013; Schmitz et al., 2008; Schoberer et al., 2019; Tu et al., 2008). Many questions remain as to the details of this degradation mechanism. It is unclear whether enzymes leaking out of the Golgi take a direct path to endosomes and hence lysosomes, or whether they first travel to the plasma membrane where they are rapidly endocytosed. In either case there must be cues for the detection of Golgi enzyme escapees. One possibility is that the

lumenal portion of the enzymes is recognised in the context of the late-Golgi or post-Golgi compartments. The apparent general nature of lysosomal-targeting makes this mode of recognition seem unlikely owing to the great sequence and structural diversity of the lumenal domains of Golgi enzymes (Moremen et al., 2012; Wang et al., 2017). In contrast, the short cytoplasmic tails and short TMDs are features conserved across Golgi enzymes and across eukaryota. Is the hydrophobic mismatch of a Golgi enzyme's short TMD in a thicker late-Golgi or post-Golgi membrane bilayer somehow sensed by a receptor? In yeast, a complex containing the Golgi-resident multipass transmembrane protein Tul1 has been implicated in vacuolar targeting of membrane proteins (Reggiori and Pelham, 2002; Tong et al., 2014). Tul1 is an E3 ubiquitin ligase that recognises membrane proteins with exposed polar residues within the bilayer and ubiquitinates their cytosolic lysines for targeting to the vacuole for degradation (Reggiori and Pelham, 2002). It is conceivable that enzymes of the Golgi which have leaked out to the TGN are recognised by Tul1 as a consequence of hydrophobic mismatch. However, despite the apparent conservation and general nature of the lysosomal targeting of mislocalised Golgi enzymes, a Tul1 relative in metazoans is missing (Reggiori and Pelham, 2002). Furthermore, many Golgi enzymes do not have any lysine residues in their cytosolic tails to serve as a substrate for ubiquitination. While it has been shown for other substrates that the N-terminal amine group can be ubiquitinated to initiate degradation (Ciechanover and Ben-Saadon, 2004), none of the known Tul1 substrates are ubiquitinated in this way (only one of which is a Golgi enzyme – Ktr1) (Tong et al., 2014).

1.7 Pathological and physiological implications of Golgi enzyme sorting

Glycosylation enzymes often act in a pathway of sequential sugar modifications and the efficiency of these pathways is thought to be dependent on the sequential compartmentalisation of these enzymes and their relative abundance. For example, the mislocalisation of a late-acting enzyme to an earlier part of the Golgi or the ER can result in the premature termination or excessive activation of glycan branching and such changes can have physiological consequences (Bard and Chia, 2016). Similarly, a change in the levels of a particular enzyme, whether it be due to changes in extracellular secretion or lysosomal-degradation, can perturb glycan patterns. For example, knockdown of GALNT3 can trigger epithelial-mesenchymal transition through aberrant glycosylation of E-cadherin (Raghu et al., 2019). Furthermore, the relocation of GALNT1 from the Golgi to the ER drives an increase in O-glycosylation which leads to activation of matrix metalloproteinase MMP14 thus promoting tumour metastasis (Nguyen et al., 2017). Moreover, single

deletions of different enzymes associated with O-linked glycosylation in *Drosophila* can have a range of different developmental phenotypes (Hagen et al., 2009).

Overexpression of *GOLPH3* is observed in a range of cancers and its oncogenic activity has been linked to the activation of AKT signalling, although the exact mechanism remains unclear (Scott et al., 2009). Several studies have shown that a variation in GOLPH3 levels can modulate the trafficking and therefore activity of its Golgi enzyme clients leading to large-scale modification of the glycosylation status of the cell which influences various oncogenic pathways (Ali et al., 2012; Isaji et al., 2014). For example, knockdown of GOLPH3 has been reported to cause a reduction in sialylation of various proteins including receptor tyrosine kinases and integrins resulting in a reduction in AKT signalling and cell motility (Isaji et al., 2014). These defects were linked to the interaction of GOLPH3 with its sialyltransferase clients and could be rescued with the overexpression of ST6GAL1. Moreover, overexpression of the *GOLPH3* ortholog in *Drosophila melanogaster* was shown to impair heparan sulphate proteoglycan synthesis which lead to aberrant hedgehog glycosylation and subsequent morphogenic defects (Chang et al., 2013).

1.8 The hunt for new coatomer adaptors

Despite the vast array of different Golgi enzymes with their variety of cytoplasmic tails and sub-Golgi localisation patterns, there are still only a few known direct or indirect interactions between the tails and the COPI coat. This is in contrast with the known diverse cargo interactions found for the distantly related clathrin/adaptor protein (AP) coats. This diversity reflects the ability of clathrin/AP coats to interact with an array of cargo-specific adaptors (McMahon and Boucrot, 2011; Robinson, 2015). In contrast, recent proteomic profiling of COPI coats generated using different coatomer subunit isoforms suggested variation in core coat components was insufficient to alter cargo specificity (Adolf et al., 2019). While the core components of the COPI coat may not influence cargo selection directly, it is possible there are unknown cargo-specific COPI adaptors that fulfil this function. GOLPH3/Vps74 proteins appear to function primarily in recycling cis/medial-residents from the late Golgi (Schmitz et al., 2008; Tu et al., 2008), but what other factors exist? It is of note in this context that unlike metazoa and fungi, plants lack an orthologue of Vps74 (Schmitz et al., 2008; Tu et al., 2008).

Since the cytoplasmic tails of these enzymes are also small, it is conceivable that any new adaptors will simultaneously interact with the lipid bilayer, as is the case for GOLPH3/Vps74. While coincidence detection by novel adaptors may pose a technical

challenge in future affinity chromatography screens, it may be a key mechanism for coupling cargo selection with sub-Golgi localisation.

Please note that sections 1.2 - 1.8 are adapted from a published review primarily written by Lawrence Welch with minor corrections and advice from Sean Munro (Welch and Munro, 2019).

1.9 Thesis aims

Over the course of more than 30 years of work, considerable progress has been made in understanding the structural assemblies of the COPI coat (Bykov et al., 2017; Dodonova et al., 2015; Suckling et al., 2015). Furthermore, the functional role of COPI vesicles in the specific retrieval of ER-resident escapees from the Golgi back to the ER or in the intra-Golgi retrograde transport of Golgi-residents is well established (Pantazopoulou and Glick, 2019). However, our understanding of the cargo diversity between different subclasses of COPI vesicle is limited. Clear distinctions have been reported between COPI vesicles which cycle within the Golgi and those which cycle from the Golgi to the ER (Donohoe et al., 2007). However, little progress has been made in exploring the diversity amongst intra-Golgi vesicles. As we begin to explore the nuances of intra-Golgi traffic, it will help unlock the puzzle of how Golgi-resident glycan-modifying enzymes are differentially distributed across the Golgi stack. As we improve our understanding of how glycosylation enzymes are trafficked, we also make strides in our understanding of the glycobiology of eukaryotic life. It is these insights that can be leveraged in myriad scientific fields and applications including developmental biology, oncology, pathology, biopharmaceuticals, immunology, virology and many more.

To address this issue, I set out to ask a few broad questions. Firstly, what defines different subclasses of COPI vesicles cycling between different Golgi cisternae? In other words, what are the different proteins responsible for the differential sorting of cargo and trafficking of these vesicles? What are the different sorting cues in the cargo themselves which enable this differentiation? As a starting point to tackle these problems, I set about trying to identify and characterise novel intra-Golgi vesicle-resident factors.

2. Materials and Methods

2.1 Materials

2.1.1 Plasmids

Listed in Table 2.1.1 are the plasmids used in this thesis

Table 2.1.1

Unique Code	Coding Region of Interest (notes)	Description	Vector (Source)
pLGW77	FAM114A2-HA-MAO	Fused at its C-terminus to a GAGAGA linker, a HA epitope tag and the TMD of monoamine oxidase as a mitochondrial targeting sequence. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW78	FAM114A1-HA-MAO		
pLGW524	FAM114A1 ALPS (124G-140S)>NUP133 ALPS (245L-267S)-HA-MAO FAM114A1 with a region containing its ALPS motif substituted for that of NUP133		
pLGW86	FAM114A2-GGGGSLVPRGSGGGGS-BioID2-HA-MAO	Fused at its C-terminus to a large flexible linker, the promiscuous biotin ligase BioID2, a GAGAGA linker, a HA epitope tag and the TMD of monoamine oxidase as a mitochondrial targeting sequence. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW87	FAM114A1-GGGGSLVPRGSGGGGS-BioID2-HA-MAO		
pLGW139	FAM114A2 WG mutant (82-AGYAGSA-88) - GGGGSLVPRGSGGGGS-BioID2-HA-MAO FAM114A2 with the tryptophan residues of the ALPS motif/'WG' repeat mutated at positions 82, 85 and 88 to alanine		
pLGW197	FAM114A1 DUF719 (125-299)-GGGGSLVPRGSGGGGS-BioID2-HA-MAO		
pLGW198	FAM114A2 DUF719 (80-247)-GGGGSLVPRGSGGGGS-BioID2-HA-MAO		
pLGW201	FAM114A1 WG mutant 127-AAGAGSA-133 - GGGGSLVPRGSGGGGS-BioID2-HA-MAO FAM114A1 with the tryptophan residues of the ALPS motif/'WG' repeat mutated at positions 127, 130 and 133 to alanine.		

pLGW90	FAM114A1-GAGA-GFP	Fused at its C-terminus to a GAGA linker and GFP. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW91	FAM114A1 (1-410V)-GAGA-GFP		
pLGW95	FAM114A2-GAGA-GFP		
pLGW128	FAM114A2 WG mutant (82-AGYAGSA-88)-GAGA-GFP FAM114A2 with the tryptophan residues of the ALPS motif/'WG' repeat mutated at positions 82, 85 and 88 to alanine		
pLGW138	FAM114A2 ΔCC (268-295)-GAGA-GFP FAM114A2 with the predicted coiled-coil between residues 268-295 deleted		
pLGW146	FAM114A1 DUF719 (125-299)-GAGA-GFP		
pLGW147	FAM114A2 DUF719 (80-247)-GAGA-GFP		
pLGW343	FAM114A2 ALPS motif (73V-133V)-GAGA-GFP A region containing the ALPS motif of FAM114A2		
pLGW344	FAM114A2 ALPS motif (73V-133V) WG mutant (82-AGYAGSA-88)-GAGA-GFP A region containing the ALPS motif of FAM114A2 with the tryptophan residues mutated at positions 82, 85 and 88 to alanine.		
pLGW140	eSpCas9(1.1) FAM114A2 Exon 2 #TACTCTCTGGTTTGGCACCT sgRNA targeting *FAM114A2* at a specific locus in exon 2.	A bicistronic CRISPR-Cas9 mammalian expression plasmid. N-terminally 3x FLAG-tagged enhanced specificity Cas9 from *S. pyogenes* fused under a CAG promoter. U6 promoter driving expression of an sgRNA.	eSpCas9(1.1) (Feng Zhang)
pLGW142	eSpCas9(1.1) FAM114A1 Exon 3 #GTGCAGGGGCTGCCGCCATT sgRNA targeting *FAM114A1* at a specific locus in exon 3.		

pLGW321	SpCas9(BB)-2A-GFP FAM114A1 Exon 3 #CCAACACCAGCTGACCCCAG sgRNA targeting *FAM114A1* at a specific locus in exon 3.	A bicistronic CRISPR-Cas9 mammalian expression plasmid. N-terminally 3x FLAG-tagged Cas9 from *S. pyogenes* fused to a T2A sequence and an EGFP tag under a CAG promoter. U6 promoter driving expression of an sgRNA	pX458 (Feng Zhang)
pLGW322	SpCas9(BB)-2A-GFP FAM114A2 Exon 2 #GGGGCTGCTTCAGTTAGCAG sgRNA targeting *FAM114A2* at a specific locus in exon 2.		
pLGW443	SpCas9(BB)-2A-GFP GOLPH3 Exon 2 #GAGAGGAAGGTTACAACTAG sgRNA targeting *GOLPH3* at a specific locus in exon 2.		
pLGW444	SpCas9(BB)-2A-GFP GOLPH3L Exon 2 #CTTCTTCCATAAGAGTAAGG sgRNA targeting *GOLPH3L* at a specific locus in exon 2.		
pLGW445	SpCas9(BB)-2A-GFP GOLPH3L Exon 2 #GTAATGCAGTTAGGTTTGCT sgRNA targeting *GOLPH3L* at a specific locus in exon 2.		
pLGW171	FAM114A1-GAGA-GST	Fused at its C-terminus to a GAGA linker and a GST tag for expression in insect cells using the MultiBac baculovirus expression system.	pACEBac1 (Geneva Biotech)
pLGW172	FAM114A2-GAGA-GST		
pLGW196	FAM114A2 WG mutant 82-AGYAGSA-88 -GAGA-GST FAM114A2 with the tryptophan residues of the ALPS motif/'WG' repeat mutated at positions 82, 85 and 88 to alanine.		
pLGW203	FAM114A1 127-AAGAGSA-133 -GAGA-GST FAM114A1 with the tryptophan residues of the ALPS motif/'WG' repeat mutated at positions 127, 130 and 133 to alanine.		
pLGW308	FAM114A2 (1-79Q)-GAGA-GST		
pLGW309	FAM114A2 (1-347P)-GAGA-GST		

pLGW310	FAM114A2 (80T-347P)-GAGA-GST	Fused at its C-terminus to a GAGA linker and a GST tag for expression in insect cells using the MultiBac baculovirus expression system.	pACEBac1 (Geneva Biotech)
pLGW311	FAM114A2 (80T-505)-GAGA-GST		
pLGW312	FAM114A2 (348L-505)-GAGA-GST		
pLGW538	FAM114A2 (348L-404R)-GAGA-GST		
pLGW539	FAM114A2 (405K-441T)-GAGA-GST		
pLGW540	FAM114A2 (442A-505)-GAGA-GST		
pLGW306	FAM114A1-GAGA-GST	Fused at its C-terminus to a GAGA linker and a GST tag for expression in bacteria.	pOPC (Olga Perisic)
pLGW307	FAM114A2-GAGA-GST		
pLGW543	CG9590 (1-332)-GSGSGS-TEV-GST in pOPC A truncated form of the *D. melanogaster* ortholog of FAM114A, CG9590 (1-332)		
pLGW608	CG9590 (1-78S)-GSGSGS-TEV-GST A truncated form of the *D. melanogaster* ortholog of FAM114A, CG9590 (1-78)		
pLGW609	CG9590 (1-78S) W24,27A - GSGSGS-TEV-GST A truncated form of the *D. melanogaster* ortholog of FAM114A, CG9590 (1-78), with alanine to tryptophan point mutations at positions 24 and 27 which are predicted to ablate to a $Wx_{n(1-6)}$[W/F] motif.		
pLGW617	CG9590 (1-332) W24,27A - GSGSGS-TEV-GST A truncated form of the *D. melanogaster* ortholog of FAM114A, CG9590 (1-78), with alanine to tryptophan point mutations at positions 24 and 27 which are predicted to ablate to a $Wx_{n(1-6)}$[W/F] motif.		
pLGW327	GST-TEV-GOLPH3	Fused at its N-terminus to a TEV protease cleavage site and a GST tag for expression in bacteria	pOPTG (Olga Perisic)
pLGW328	GST-TEV-GOLPH3L		
pLGW544	GST-TEV-CG9590 (1-332) A truncated form of the *D. melanogaster* ortholog of FAM114A, CG9590 (1-332)		

pLGW435	GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2	Fused to the TMD and short lumenal section of sucrase-isomaltase, a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW449	B3GALT6 (1-11R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of B3GALT6		
pLGW451	ST6GAL1 (1-9K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of ST6GAL1		
pLGW455	CHPF2 (1-6L)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of CHPF2		
pLGW457	HS2ST1 (1-11K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of HS2ST1		
pLGW459	GOLIM4 (1-12R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GOLIM4		
pLGW460	MGAT2 (1-9K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of MGAT2		
pLGW485	B3GAT3 (1-7N)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of B3GAT3		
pLGW486	GOLM1(1-12K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GOLM1		
pLGW487	GALNT3 (1-19K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT3		
pLGW488	CHSY1 (1-7R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of CHSY1		

pLGW489	UXS1 (1-19K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of UXS1	Fused to the TMD and short lumenal section of sucrase-isomaltase, a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW490	TPST2 (1-8V)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of TPST2		
pLGW491	MANEAL (1-8R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of MANEAL		
pLGW492	GALNT12 (1-19E)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT12		
pLGW493	CASC4 (1-14P)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of CASC4		
pLGW494	MGAT5 (1-13K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of MGAT5		
pLGW507	GALNT2 R2K (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with an arginine to lysine point mutation at position 2		
pLGW508	GALNT2 R3K (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with an arginine to lysine point mutation at position 3		
pLGW509	GALNT2 R4K (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with an arginine to lysine point mutation at position 4		
pLGW510	GALNT2 RRR234KKK (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with arginine to		

	lysine point mutations at positions 2, 3 and 4	Fused to the TMD and short lumenal section of sucrase-isomaltase, a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW512	B3GALT6 RRR9,10,11KKK (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of B3GALT6 with arginine to lysine point mutations at positions 9, 10 and 11		
pLGW513	MANEAL R5K (1-8R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of MANEAL with an arginine to lysine point mutation at position 5		
pLGW514	MANEAL R5A (1-8R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of MANEAL with an arginine to alanine point mutation at position 5		
pLGW515	GCNT1 (1-9R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GCNT1		
pLGW534	GCNT1 R7H (1-9R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GCNT1 with an arginine to histidine point mutation at position 7		
pLGW535	GCNT1 R7A (1-9R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GCNT1 with an arginine to alanine point mutation at position 7		
pLGW536	GALNT2 (1-6R)-AA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal double alanine insertion		
pLGW537	GALNT2 (1-6R)-AAA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal triple alanine insertion		

pLGW550	FUT3 (1-15R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of FUT3	Fused to the TMD and short lumenal section of sucrase-isomaltase, a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW551	GALNT4 (1-12C)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT4		
pLGW555	TPST1 (1-8N)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of TPST1		
pLGW556	GALNT2 (1-6R)-AAAA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal quadruple alanine insertion		
pLGW557	GALNT2 (1-6R)-AAAAA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal quintuple alanine insertion		
pLGW558	GALNT2 (1-6R)-AAAAAA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal sextuple alanine insertion		
pLGW559	GALNT2 (1-6R)-AAAAAAA-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 with a membrane-proximal septuple alanine insertion		
pLGW563	EXTL3 (1-11T)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of EXTL3		

pLGW452	SI (1-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase.	N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail, TMD and a short lumenal section fused to a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW530	SI (1-12S)-RRR-SI (13L-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase with a membrane-proximal triple arginine insertion in its cytoplasmic tail		
pLGW531	SI (1-12S)-KKK-SI (13L-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a membrane-proximal triple lysine insertion in its cytoplasmic tail		
pLGW532	SI (1-12S)-HHH-SI (13L-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a membrane-proximal triple histidine insertion in its cytoplasmic tail		
pLGW533	SI (1-10E)-RIRSR-SI (13L-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a with a triple arginine residue intercalation in its cytoplasmic tail		
pLGW571	SI S7D (1-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a serine to aspartate point mutation at position 7 in the cytoplasmic tail		
pLGW572	SI S7A (1-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a serine to alanine point mutation at position 7 in the cytoplasmic tail		
pLGW573	SI E10A (1-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a glutamate to alanine point mutation at position 10 in the cytoplasmic tail		
pLGW574	SI S7A, E10A (1-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with alanine point mutations at		

	positions 7 and 10 in the cytoplasmic tail	N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail, TMD and a short lumenal section fused to a GAGA linker, a GFP tag and a FLAG tag. For expression in mammalian cells.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW575	SI F6A (1-38V)-GAGA-GFP-FLAG Sucrase-isomaltase with a phenylalanine to alanine point mutation at position 6 in the cytoplasmic tail.		
pLGW580	GST-GOLPH3 Δ1-51 GOLPH3 with the N-terminal region 1-51 deleted	fused to a thrombin cleavage site and a GST tag at its N-terminus. For expression in bacteria.	pGEX-4T2 (GE Life Sciences)
pLGW582	GST-GOLPH3 RR14,15AA GOLPH3, with two conserved arginine residues mutated to alanines at positions 14 and 15 reported to ablate coatomer binding (Tu et al., 2012).		
pLGW606	6xHis-TEV-MHD DmδCOP (291-532) The μ-homology domain of the *D. melanogaster* δ-COP subunit.	fused to a TEV protease site and 6xHis tag at its N-terminus. For expression in bacteria.	pOPTH (Olga Perisic)
pLGW607	6xHis-TEV-MHD DmδCOP HAKS mutant (291-532) H350A, K355S The μ-homology domain of the *D. melanogaster* δ-COP subunit, with point mutations in key residues predicted to ablate binding to $Wx_{n(1-6)}$[W/F] motif-containing proteins based on sequence homology from a previous report (Suckling et al., 2015).		
pLGW619	BFP-P2A-T2A-GOLPH3	downstream of a BFP tag and a tandem P2A-T2A self-cleaving peptide sequence. For expression in mammalian cells	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW620	BFP-P2A-T2A-GOLPH3 RR14,15AA GOLPH3, with two conserved arginine residues mutated to alanines at position 14 and 15 reported to ablate coatomer binding (Tu et al., 2012).		
pLGW675	GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of GALNT2 fused to the TMD and short lumenal section of sucrase-isomaltase	Fused to a GAGA linker, a GFP tag and a FLAG tag. For the generation of cumate-inducible stable mammalian cell lines using the PiggyBac transposase system.	pJJS359 (John Shin)
pLGW676	SI (1-38V)-GAGA-GFP-FLAG		

	N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail, TMD and a short lumenal section	Fused to a GAGA linker, a GFP tag and a FLAG tag. For the generation of cumate-inducible stable mammalian cell lines using the PiggyBac transposase system.	pJJS359 (John Shin)
pLGW690	SI (1-12S)-GALNT2 (7M-24Y)-SI (33T-38V)-GAGA-GFP-FLAG-P2A-T2A-mCherry N-terminal region of sucrase-isomaltase but the TMD substituted for that of GALNT2		
pLGW691	SI (1-12S)-ST6GAL1 (10F-26W)-SI (33T-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase but the TMD substituted for that of ST6GAL1		
pLGW692	ST6GAL1 (1-9K)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of ST6GAL1 fused to the TMD and short lumenal section of sucrase-isomaltase		
pLGW693	TPST1 (1-8N)-SI (13L-38V)-GAGA-GFP-FLAG N-terminal cytoplasmic tail of TPST1 fused to the TMD and short lumenal section of sucrase-isomaltase		
pLGW694	SI (1-12S)-RRR-SI (13L-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail with a membrane-proximal triple arginine insertion		
pLGW695	SI (1-12S)-KKK-SI (13L-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail with a membrane-proximal triple lysine insertion		
pLGW696	SI (1-12S)-HHH-SI (13L-38V)-GAGA-GFP-FLAG N-terminal region of sucrase-isomaltase consisting of the cytoplasmic tail with a		

	membrane-proximal triple histidine insertion		
pLGW462	Str-KDEL IRES> GALNT2 CTS (1-112L)-SBP-EGFP	A Bicistronic RUSH plasmid encoding an ER hook consisting of the signal peptide of STIM1 fused to core streptavidin and a KDEL sequence under a CMV promoter. Also encoding a downstream IRES sequence and the CTS domain of GALNT2 fused to streptavidin binding protein and EGFP.	RUSH1 (Frank Perez)
pLGW733	GALNT2 (1-112L)-GAGA-GFP-FLAG-P2A-T2A-mCherry	CTS domain of GALNT2 upstream of a tandem P2A-T2A self-cleaving peptide sequence and mCherry. For the generation of stable mammalian cell lines using the PiggyBac transposase system.	PiggyBac-compatible pcDNA3.1+
pLGW98	GFP	GFP, for expression in mammalian cells. Used as a single colour control for compensation purposes in flow cytometry assays.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW462	MGAT2 (1-9K)-SI (13L-38V)-GAGA-mCherry-KDEL	mCherry-tagged reporter for expression in mammalian cells. Used as a single colour control for compensation purposes in flow cytometry assays.	pcDNA3.1+ (Thermo Fischer Scientific)
pLGW173	GST	GST, for expression in insect cells using the MultiBac baculovirus expression system. Used as a negative control for GST affinity chromatography.	pACEBac1 (Geneva Biotech)
pOPTG	GST	GST, for expression in bacteria. Used as a negative control for GST affinity chromatography.	pOPTG (Olga Perisic)
pJJS345	BioID2-HA-MAO	BioID2 fused at its C-terminus to a GAGAGA linker, a HA epitope tag and the TMD of monoamine oxidase as a mitochondrial targeting sequence. Used as a negative control for mitochondrial relocation assays. For expression in mammalian cells.	pcDNA3.1+ (John Shin, Thermo Fischer Scientific)
PB210PA-1	Super PiggyBac Transposase	Co-transfected with PiggyBac-compatible vectors to promote genomic integration.	(Systems Biosciences)
pCFD3_1, 2,3,4	sgRNAs targeting 5' or 3' UTR of *CG9590*.	Expression of sgRNA under a U6 promoter. For injection and expression in *Drosophila*.	pCFD3 (Nadine Muschalik, Simon Bullock)

2.1.2 Cell Lines and Strains

Listed in Table 2.1.2.1 are the cell lines used in this thesis

Table 2.1.2.1

Unique Identifier	Description	Origin/Parental Line
D.Mel-2	*Drosophila Melanogaster* cell line derived from embryonic tissue.	ATCC
HeLa	Human cell line derived from a cervical adenocarcinoma cell	ATCC
HEK293T	Human cell line derived from an embryonic kidney. Constitutively expressing SV40 large T antigen.	ATCC
Sf9	*Spodoptera frugiperda* cell line cloned from the parental cell line IPLB-Sf-21-AE	Jianguo Shi (MRC LMB Baculovirus Facility)
U2OS	Human cell line derived from an osteosarcoma cell	ATCC
pJJS138 HEK293	Doxycycline-inducible stable cell line expressing BirA-GSGSGS-GCC185-HA-MAO. Generated through integration of pJJS138 through Flp recombination.	John Shin
pJJS143 HEK293	Doxycycline-inducible stable cell line expressing golgin-84 (1-38)-GGGGSLVPRGSGGGGG-BirA-GCC185-HA-MAO. Generated through integration of pJJS143 through Flp recombination.	John Shin
pJJS144 HEK293	Doxycycline-inducible stable cell line expressing GMAP210 (1-38)-GGGGSLVPRGSGGGGG-BirA-GCC185-HA-MAO. Generated through integration of pJJS144 through Flp recombination.	John Shin
pJJS146 HEK293	Doxycycline-inducible stable cell line expressing golgin-84 W3A (1-38)-GGGGSLVPRGSGGGGG-BirA-GCC185-HA-MAO. Generated through integration of pJJS146 through Flp recombination.	John Shin
pJJS147 HEK293	Doxycycline-inducible stable cell line expressing GMAP210 W4A (1-38)-GGGGSLVPRGSGGGGG-BirA-GCC185-HA-MAO. Generated through integration of pJJS147 through Flp recombination.	John Shin
Δ*FAM114A2* #5.5 U2OS	*FAM114A2* deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW140 and pLGW322.	U2OS
Δ*FAM114A1* #6.2 U2OS	*FAM114A1* deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW142 and pLGW321.	U2OS
ΔΔ*FAM114A1+2* #7.18 U2OS	*FAM114A1, FAM114A2* double deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW321 and pLGW322.	U2OS

ΔΔ*GOLPH3+3L* #8.1 U2OS	*GOLPH3, GOLPH3L* double deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW443, pLGW444 and pLGW445.	U2OS
ΔΔ*GOLPH3+3L* #8.5 U2OS	*GOLPH3, GOLPH3L* double deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW443, pLGW444 and pLGW445.	U2OS
ΔΔΔΔ*FAM114A1+2, GOLPH3+3L* U2OS	*FAM114A1, FAM114A2, GOLPH3, GOLPH3L* quadruple deletion mutant U2OS cell line. Deletion introduced through CRISPR-Cas9 gene-editing using pLGW443, pLGW444 and pLGW445 in a *ΔΔFAM114A1+2* U2OS background.	*ΔΔFAM114A1+2* #7.18 U2OS
pLGW675 U2OS	Polyclonal cumate-inducible stable cell line expressing GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW675 using PiggyBac transposase.	U2OS
pLGW676 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW676 using PiggyBac transposase.	U2OS
pLGW690 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-GALNT2 (7M-24Y)-SI (33T-38V)-GAGA-GFP-FLAG-P2A-T2A-mCherry in a wild-type U2OS background. Generated through integration of pLGW690 using PiggyBac transposase.	U2OS
pLGW691 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-ST6GAL1 (10F-26W)-SI (33T-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW691 using PiggyBac transposase.	U2OS
pLGW692 U2OS	Polyclonal cumate-inducible stable cell line expressing ST6GAL1 (1-9K)-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW692 using PiggyBac transposase.	U2OS
pLGW693 U2OS	Polyclonal cumate-inducible stable cell line expressing TPST1 (1-8N)-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW693 using PiggyBac transposase.	U2OS
pLGW694 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-RRR-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW694 using PiggyBac transposase.	U2OS
pLGW695 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-KKK-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW695 using PiggyBac transposase.	U2OS

pLGW696 U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-HHH-SI (13L-38V)-GAGA-GFP-FLAG in a wild-type U2OS background. Generated through integration of pLGW696 using PiggyBac transposase.	U2OS
pLGW733 U2OS	Stable cell line expressing GALNT2 (1-112L)-GAGA-GFP-FLAG-P2A-T2A-mCherry in a wild-type U2OS background. Generated through integration of pLGW733 using PiggyBac transposase.	U2OS
pLGW675 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW675 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW676 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW676 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW690 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-GALNT2 (7M-24Y)-SI (33T-38V)-GAGA-GFP-FLAG-P2A-T2A-mCherry in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW690 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW691 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-ST6GAL1 (10F-26W)-SI (33T-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW691 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW692 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing ST6GAL1 (1-9K)-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW692 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW693 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing TPST1 (1-8N)-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW693 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW694 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-RRR-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW694 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW695 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-KKK-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW695 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS

pLGW696 *ΔΔGOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (1-12S)-HHH-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔGOLPH3, GOLPH3L* U2OS background. Generated through integration of pLGW696 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW733 *ΔΔGOLPH3+3L* U2OS	Stable cell line expressing GALNT2 (1-112L)-GAGA-GFP-FLAG-P2A-T2A-mCherry in a *ΔΔGOLPH3+3L* U2OS background. Generated through integration of pLGW733 using PiggyBac transposase.	*ΔΔGOLPH3+3L* #8.1 U2OS
pLGW675 *ΔΔFAM114A1+2,* U2OS	Polyclonal cumate-inducible stable cell line expressing GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔFAM114A1+2* U2OS background. Generated through integration of pLGW675 using PiggyBac transposase.	*ΔΔFAM114A1+2* #7.18 U2OS
pLGW676 *ΔΔFAM114A1+2,* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔFAM114A1+2* U2OS background. Generated through integration of pLGW676 using PiggyBac transposase.	*ΔΔFAM114A1+2* #7.18 U2OS
pLGW675 *ΔΔΔΔFAM114A1+2, GOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing GALNT2 (1-6R)-SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔΔΔFAM114A1+2, GOLPH3+3L* U2OS background. Generated through integration of pLGW675 using PiggyBac transposase.	*ΔΔΔΔFAM114A1+2, GOLPH3+3L* #7.18 U2OS
pLGW676 *ΔΔΔΔFAM114A1+2, GOLPH3+3L* U2OS	Polyclonal cumate-inducible stable cell line expressing SI (13L-38V)-GAGA-GFP-FLAG in a *ΔΔΔΔFAM114A1+2, GOLPH3+3L* U2OS background. Generated through integration of pLGW676 using PiggyBac transposase.	*ΔΔΔΔFAM114A1+2, GOLPH3+3L* #7.18 U2OS

Listed in Table 2.1.2.2 are the bacterial strains used in this thesis.

Table 2.1.2.2

Strain	**Supplier**	**Genotype**
α-Select Silver Efficiency	Bioline	F^- *deoR endA1 recA1 relA1 gyrA96 hsdR17*(r_k^- , m_k^+) *supE44 thi*-1 *pho*A Δ(*lac*ZYA *arg*F)U169 Φ80lacZΔM15λ^-
NEB 5-alpha Competent *E. coli* (High efficiency)	New England Biolabs	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
BL21-CodonPlus(DE3)-RIL competent cells	Agilent	*E. coli* B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^r]
DH10EMBacY	Geneva Biotech	(Consult supplier's manual)

2.1.3 Antibodies and Lectins

Listed in Table 2.1.3.1 are the primary antibodies used in this thesis.

Table 2.1.3.1

Antigen	Species	Dilution		Supplier (clone/catalogue #)
		WB	IF	
6xHis	Mouse	1:1000		Abcam (ab18184)
β-COP	Rabbit	1:1000		Abcam (ab2899)
ERGIC-53	Rabbit		1:100	Santa Cruz Biotech (H-245/sc-66880)
FAM114A1	Rabbit	1:1000		Atlas Antibodies (HPA069701, HPA069104 used in tandem)
FAM114A2	Rabbit	1:1000		Atlas Antibodies (HPA035870)
FLAG	Mouse	1:2000		Sigma (M2)
GALNT2	Rabbit		1:50	Abcam (ab102650)
GALNT4	Sheep	1:3000		R&D Systems (AF7528)
Giantin	Goat		1:100	Santa Cruz Biotech (N-18/sc-46993)
GM130	Mouse		1:300	BD Biosciences (610823)
Golgin-245	Mouse		1:300	BD Biosciences (611281)
Golgin-84	Rabbit		1:300	Atlas Antibodies (HPA000992)
GOLPH3	Rabbit	1:2000	1:50	Protein Tech (19112-1-AP)
HA	Rat		1:300	Roche (3F10)
TGN46	Sheep		1:300	ABD serotec (AHP500G)
Tubulin	Rat	1:250		In house (YL1/2)
GFP*	Rat			Biolegend (FM264G)

*Conjugated to Alexa Fluor 647 and used for flow cytometry assays at a dilution of 1:20.

Listed in Table 2.1.3.2 are the secondary antibodies used in this thesis.

Table 2.1.3.2

Specificity Antigen	Species	Dilution	Supplier (clone/catalogue #)
Anti-Rabbit Alexa Fluor 488	Donkey	1:300	Thermo Fischer Scientific (A21206)
Anti-Rabbit Alexa Fluor 555	Donkey	1:300	Thermo Fischer Scientific (A31572)
Anti-Rabbit Alexa Fluor 647	Donkey	1:300	Thermo Fischer Scientific (A31573)
Anti-Rabbit HRP	Donkey	1:3000	GE Life Sciences (NA934V)
Anti-Mouse Alexa Fluor 647	Donkey	1:300	Thermo Fischer Scientific (A31571)
Anti-Mouse HRP	Sheep	1:3000	GE Life Sciences (NA931V)
Anti-Rat Alexa Fluor 488	Donkey	1:300	Thermo Fischer Scientific (A21208)
Anti-Rat Alexa Fluor 555	Donkey	1:300	Abcam (ab150154)
Anti-Rat HRP	Goat	1:3000	Abcam (ab205720)
Anti-Goat Alexa Fluor 555	Donkey	1:300	Thermo Fischer Scientific (A21432)
Anti-Sheep Alexa Fluor 555	Donkey	1:300	Thermo Fischer Scientific (A21436)
Anti-Sheep Alexa Fluor 647	Donkey	1:300	Thermo Fischer Scientific (A21448)
Anti-Sheep HRP	Donkey	1:3000	Santa Cruz Biotech (sc-2916)

Listed in Table 2.1.3.3 are the lectins and their corresponding inhibitory sugars that were used in this thesis.

Table 2.1.3.3

Lectin Name	Specificity (Inhibiting Sugar)
Datura stramonium lectin (DSL)	[GlcNAc]1-3, *N*-acetylglucosamine (1:20, chitin hydrolysate)
Erythrina cristagalli lectin (ECL)	Galactose, *N*-acetylgalactosamine, lactose (200 mM lactose)
Jacalin (Jac)	T-antigen/core 1 O-glycan/galactosyl (β-1,3) *N*-acetylgalactosamine (800 mM galactose)
Griffonia simplicifolia lectin II (GSLII)	α- or β-linked *N*-acetylglucosamine (1:20, chitin hydrolysate)
Lycopersicon esculentum lectin (LEL)	[GlcNAc]1-3, *N*-Acetylglucosamine (1:20, chitin hydrolysate)
Solanum tuberosum lectin (STL)	*N*-acetylglucosamine (1:20, chitin hydrolysate)
Vicia villosa lectin (VVA)	Tn-antigen/a single O-linked α-*N*-acetylgalactosamine residue (200 mM *N*-acetylgalactosamine)
Wisteria floribunda lectin (WFA)	*N*-acetylgalactosamine (200 mM *N*-acetylgalactosamine)

All lectins were conjugated to fluorescein while biotinylated forms of Jac and VVA were also used. All lectins were sourced from Vector Biolabs and used at a final concentration of 20 µg/ml. All inhibiting sugars were sourced from Sigma with the exception of chitin hydrolysate (Vector Biolabs)

2.2 Methods

2.2.1 Molecular Cloning

2.2.1.1 Transformation, Growth and Induction of *Escherichia Coli*

Aliquots of bacterial strains were stored at -80°C and were thawed on ice for 30 minutes prior to transformation. Plasmid DNA was added at a maximum volume ratio of 1:10 to the bacterial aliquot volume and mixed briefly and gently with a pipette tip on ice. The mixture was incubated on ice for 30 minutes, subjected to heat shock at 42°C for 45 seconds and briefly returned to ice. SOB medium was added to BL21 and DH10EMBacY strains and cells were incubated at 37°C with agitation to either allow for sufficient outgrowth (at least 1 hour, BL21) or for transposition of pACEBac1 vectors into the baculoviral genome (4 hours, DH10EMBacY) according to the manufacturer's protocol (Geneva Biotech). Transformed bacteria was plated on 2x TY or LB medium agar plates supplemented with selective antibiotics and grown overnight at 37°C. See Table 2.2.1.1 for the media recipes and antibiotics used for bacterial growth in this thesis. Resulting single colonies were used for the inoculation of liquid cultures of 2x TY medium supplemented with selective antibiotics. Successfully transposed DH10EMBacY colonies were also screened for based on blue-white selection. Bacterial plates were kept at 4°C for up to a month as necessary. After a month, strains that were still required were restreaked on fresh selective plates and grown and stored as described above.

Single colonies of BL21 clones expressing GST-tagged fusion proteins were used to inoculate overnight liquid starter cultures which were grown at 37°C, 220 rpm. Starter cultures were used to inoculate larger cultures at a ratio of 1:25 and were incubated until they reached an OD_{600} of 0.5-0.8. Cultures were induced with 100 µM IPTG (Melford) 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 phosphate-buffered saline (PBS) by resuspension and centrifugation. Cells were kept on ice prior to immediate use or snap frozen in liquid nitrogen and kept at -80°C for long term storage.

Table 2.2.1.1

Reagent	Description	Source
2x TY agar plate	1% tryptone, 0.5% yeast extract, 0.8% NaCl, 1.5% agar in H_2O	MRC LMB media kitchen
Bacmid LB agar plate	1% tryptone, 0.5% yeast extract, 0.8% NaCl, 1.5% agar in H_2O kanamycin 50 µg/ml, tetracycline 10 µg/ml, gentamycin 10 µg/ml, bluo-gal 100 µg/ml, IPTG 40 µg/ml	MRC LMB media kitchen
2x TY medium	1.6% tryptone, 1% yeast extract, 0.5% NaCl in H_2O	MRC LMB media kitchen
SOB medium	2% tryptone, 0.5% yeast extract, 0.5% NaCl in H_2O	MRC LMB media kitchen
Ampicillin	Used at a final concentration of 100 µg/ml for transformed DH5-α and BL21 strains	Alfa Aesar
Chloramphenicol	Used at a final concentration of 25 µg/ml in conjunction with ampicillin for transformed BL21 strains.	Formedium
G418	Used at a final concentration of 100 µg/ml for liquid cultures of DH10EMBacY strains	Alfa Aesar

Percentages are given as weight/volume.

2.2.1.2 DNA Manipulation

DNA manipulations conducted using specified kits were carried out according to the manufacturer's protocol unless stated otherwise. Plasmid DNA was prepared from bacterial liquid cultures using Qiagen plasmid kits. Bacmid DNA was purified from bacterial liquid cultures by isopropanol precipitation according the MultiBac manual (Geneva Biotech). Mammalian genomic DNA was purified from cells in culture using the Qiagen Puregene DNA extraction kit. Both genomic and plasmid DNA were eluted in TE buffer. DNA was amplified by PCR using Phusion High-Fidelity DNA Polymerase (New England Biolabs) with GC buffer, an annealing time of 20 seconds and an extension time of 30 s/kbp. Plasmid and genomic DNA templates were used in PCR reactions at a final concentration of 0.5 and 2 ng/µl respectively. Plasmid DNA and PCR fragments were subjected to restriction enzyme digestion for 90 minutes at 37°C in CutSmart buffer (New England Biolabs) followed by heat inactivation at 80°C for 20 minutes. All

restriction enzymes were supplied by New England Biolabs. PCR products (prior to digestion) and digested plasmids were resolved by gel electrophoresis using a 0.8-2% agarose gel in the presence of SYBR Safe DNA gel stain (Thermo Fischer Scientific). A 1 Kb Plus DNA ladder (Thermo Fischer Scientific) was resolved alongside PCR and digestion products and gels were visualised using a Safe Image Blue-Light Transilluminator (Thermo Fischer Scientific). DNA bands of the correct size were excised and purified using the Qiagen gel extraction kit. Ligation reactions between DNA fragments with compatible sticky-ends were conducted using T4 DNA Ligase (New England Biolabs) for 15 minutes at room temperature.

2.2.1.3 Cloning Strategies

The N-terminal vesicle binding regions of Intra-Golgi golgins were fused to a flexible linker, the promiscuous biotin ligase BirA*, to the large coiled-coil golgin GCC185 with its Golgi targeting GRIP domain replaced with the mitochondrial-targeting sequence of monoamine oxidase (MAO, 481-527) and a HA epitope tag. These chimeric fusions where cloned into pcDNA5/FRT/TO for the generation of doxycycline-inducible HEK293 cell lines using the Flp-In T-REx system (Thermo Fischer Scientific). All of this cloning was conducted by John Shin.

FAM114A proteins and associated truncation mutants were amplified by PCR from existing FAM114A plasmids (Allison Gillingham). The same templates were used to generate FAM114A ALPS/'WG' mutants by *in vivo* assembly (IVA; García-Nafría et al., 2016) in which tryptophan residues at positions 127, 130 and 133 for FAM114A1 and positions 82, 85 and 88 for FAM114A2 were simultaneously mutated to alanines. A FAM114A2 mutant with the deletion of the predicted coiled-coil between residues 268-295 was also generated by IVA. A chimera, in which a region containing the predicted ALPS motif of FAM114A1 (124-140) was replaced with the ALPS motif of the functionally unrelated NUP133 (245-267), was generated through gBlock synthesis (IDT) and restriction enzyme cloning. For mitochondrial relocation assays, FAM114A proteins were fused at their C-terminus to either an HA tag and MAO sequence or a large flexible linker, the promiscuous biotin ligase BioID2, a HA tag and MAO sequence using the restriction sites NheI and KpnI in the vector pcDNA3.1+. FAM114A proteins used for immunofluorescence experiments were fused to a GAGA linker and GFP at their C-terminus using the restriction sites NheI and KpnI in pcDNA3.1+.

Plasmids were designed for the baculoviral or bacterial expression and subsequent purification of GST-tagged GOLPH3 and FAM114A family proteins. GOLPH3, GOLPH3L, a GOLPH3 mutant with two conserved arginine residues mutated to alanines at positions 14 and 15 reported to ablate coatomer binding (Tu et al., 2012) and a truncated form of CG9590 (1-332) were synthesised as gBlocks. GOLPH3 was codon optimised to reduce its GC content in the N-terminus to facilitate synthesis. A mutant CG9590 with simultaneous tryptophan to alanine point

mutants at positions 24 and 27, which was predicted to ablate binding to a $Wx_{n(1-6)}$[W/F] δ-COP-binding motif, was generated by IVA. GOLPH3 and GOLPH3L family proteins were 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. Mutant GOLPH3 RR14,15AA and an N-terminal deletion (Δ1-51) mutant GOLPH3 were fused at their N-terminus to a GST tag, a thrombin cleavage site and a 6xHis tag in pGEX-4T2 using the restriction sites BamHI and NotI. FAM114A proteins and associated truncations were fused at their C-terminus to a GAGA linker and a GST tag and cloned into either bacterial expression vector pOPC (Olga Perisic) using the restriction sites NdeI and NotI or baculoviral expression vector pACEBac1 using the sites BamHI and XbaI. CG9590 truncations were fused at their C-terminus to a GSGSGS linker, a TEV protease cleavage site and a GST tag and cloned into vector pOPC using the restriction sites NdeI and BamHI for bacterial expression. The μ-homology domain of the *Drosophila melanogaster* ortholog of δ-COP (291-532) and a mutant form with point mutations in key residues (H350A, K355S) predicted to ablate binding to $Wx_{n(1-6)}$[W/F] motif-containing proteins (Suckling et al., 2015) were synthesised as gBlocks. They were fused at their N-terminus to a TEV protease cleavage site and a 6xHis tag in the vector pOPTH for bacterial expression using the restriction sites NdeI and BamHI.

Plasmids were designed for the transient mammalian expression of Golgi enzyme chimeric GFP fusions to serve as baits to test *in vitro* binding to GOLPH3 and FAM114A proteins. The short cytoplasmic tail and TMD of the type II transmembrane protein sucrase-isomaltase fused to GFP has been used previously as a 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 cloned out of genomic DNA purified from HeLa cells by PCR. The analogous N-terminal CTS domains of GALNT2 (1-112) and ST6GAL1 (1-62) were synthesised as gBlocks. These N-terminal fragments were fused at their C-terminus to a GAGA linker, a GFP tag and a FLAG tag using a KpnI restriction site. The resulting chimeric fusions were cloned into pcDNA3.1+ using the restriction sites NheI and NotI to generate transient expression plasmids. 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 in using the restriction sites NheI and KpnI. Similarly, the TMD of sucrase-isomaltase was replaced with that of GALNT2 and ST6GAL1 through gBlock synthesis and restriction enzyme cloning. The CTS domain of GALNT2 was also cloned into the RUSH1 bicistronic plasmid using the restriction sites AscI and EcoRI.

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). For the purpose of the lysosomal degradation assay, a reporter consisting of the CTS domain of GALNT2, fused at its C-terminus to a GAGA linker, a GFP tag, a FLAG tag and a downstream tandem P2A-T2A self-cleaving peptide sequence followed by mCherry was synthesised as a gBlock. Using the restriction site NheI and NotI, this was 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 FAM114A family genes in mammalian cells using CRISPR-Cas9 gene editing. Oligonucleotides pairs encoding guide RNAs (sgRNAs) targeting specific loci in FAM114A and GOLPH3 genes 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 vectors eSpCas9(1.1) or pX458 (both Feng Zhang) which encode various tagged forms of Cas9 from *S. pyogenes* under a CAG promoter and a U6 promoter driving expression of the sgRNA. For genetic rescue of *ΔΔGOLPH3, GOLPH3L* U2OS cells, GOLPH3 and GOLPH3 RR14,15AA were placed downstream of a BFP tag and a tandem P2A-T2A self-cleaving peptide sequence in the vector pcDNA3.1+ for transient mammalian expression using the restriction sites BamHI and NotI.

2.2.2 Mammalian Cell Biology

2.2.2.1 Cell Culture, Cryopreservation and Transfection

U2OS, HEK293T and HeLa cells (all ATCC) were maintained in culture medium consisting of Dulbecco's modified Eagle's medium (DMEM, Thermo Fischer Scientific) with 10% fetal bovine serum (FBS, Thermo Fischer Scientific) and penicillin-streptomycin (pen-strep, MRC LMB Media Kitchen) in a humidified incubator at 37°C with 5% CO_2. Flp-In T-Rex HEK293 cell lines expressing mitochondrial-relocated golgin-BirA* fusions were maintained in DMEM with 10% FBS, pen-strep, 5 µg/ml blasticidin (Generon) and 100 µg/ml hygromycin (Life Technologies). Cells were passaged at a ratio of 1:10 every 3-4 days in which cells were washed once with EDTA solution (MRC LMB Media Kitchen), incubated in trypsin solution (MRC LMB Media Kitchen) for 2 minutes at 37°C and subsequently resuspended in culture medium prior to replating. For long term storage, cell lines were trypsinated and resuspended in culture medium as above and the cell count calculated using the Countess II (Thermo Fischer Scientific). Cells were pelleted by centrifugation at 300 x g for 5 minutes and were resuspended in Bambanker

freezing medium (GC Lymphotec) at a density of approximately $3x10^6$ cells/ml. Cell suspensions were transferred to cryovials for storage at -80°C.

For cell line generation or immunofluorescence experiments, U2OS or HeLa cells were seeded at a density of $2x10^4$ cells/cm^2 in 6-well plates 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 0.7-2 µg DNA using Fugene 6 transfection reagent (Promega) according to the manufacturer's protocol. For immunofluorescence experiments (see 2.2.5), cells were trypsinated prior to seeding onto microscope slides at least 8 hours after transfection. For biochemistry experiments (see 2.2.6), HEK293T cells were seeded in T-75 or T-175 flasks in culture medium in a humidified incubator at 37°C with 5% CO_2. 45 or 75 µl polyethyleneimine (PEI) transfection reagent (MRC LMB Media Kitchen) was mixed with 800 or 1400 µl of Opti-MEM (Thermo Fischer Scientific) per flask and the mixture was incubated at room temperature for 5 minutes. 15 or 25 µg of DNA was mixed with the Opti-MEM-PEI mixture and incubated for a further 15 minutes at room temperature. The DNA complexes were added to the flasks and mixed briefly and gently with culture medium using figure of eight motions. 48-72 hours after transfection, cells required for biochemistry were lifted into suspension by vigorous tapping of the flask followed by an EDTA solution wash to collect residual cells. Cell suspensions were pelleted by centrifugation at 300 x g for 5 minutes at 4°C and were washed once with ice cold PBS by resuspension and centrifugation. Cells were kept on ice prior to immediate use or snap frozen in liquid nitrogen and kept at -80°C for long term storage.

2.2.2.2 Cell Line Generation

2.2.2.2.1 Doxycycline-Inducible Cell Lines

Doxycycline-inducible HEK293 cell lines expressing mitochondrial-relocated N-terminal intra-Golgi golgin-BirA* chimeras were generated by John Shin using the Flp-In T-REx system.

2.2.2.2.2 CRISPR-Cas9 Gene-Edited Cells Lines

CRISPR-Cas9 technology was used to knock-out genes of interest through the induction of frame-shift mutations and subsequent introduction of premature stop codons in early constitutive exons. Genes were either targeted at one site in order to induce small indel mutations or at two sites approximately 500 bp apart to introduce larger out-of-frame deletion mutants. sgRNAs targeting specific sites were selected according to their predicted high specificity and efficiency for the locus of interest (Doench et al., 2016; Moreno-Mateos et al., 2015; Hsu et al., 2013).

U2OS cells were transfected with bicistronic plasmids for the simultaneous expression of specific sgRNAs and Cas9-T2A-EGFP. 24 hours after transfection, GFP positive cells were sorted to one

cell per well into 96-well plates containing fresh culture medium and clones were gradually expanded to 6-well format over the course of several weeks. Clones were initially validated by genotypic screening in which regions approximately 100-200 bp either side of the predicted Cas9 cut site were amplified by PCR from purified genomic DNA. PCR fragments were analysed by Sanger sequencing in order to confirm successful gene-editing (Genewiz). Whole cells lysates of clones were analysed by western blot to confirm the absence of the protein of interest. Where possible, antibodies raised against the C-terminal portion of the protein of interest were used to detect and avoid clones expressing N-terminal truncation mutants which may exhibit residual protein activity. True knockout clones were immediately expanded and cryopreserved for long-term storage and subsequent downstream analysis.

2.2.2.2.3 PiggyBac Transposon-Modified Cell Lines

The PiggyBac Transposon system (System Biosciences) was used to generate stable cell lines expressing GFP-tagged Golgi enzyme chimeric fusion reporters under a cumate-inducible promoter or a CMV promoter. Once wild-type; *ΔΔFAM114A1, FAM114A2*; *ΔΔGOLPH3, GOLPH3L* and *ΔΔΔΔFAM114A1, FAM114A2, GOLPH3, GOLPH3L* CRISPR knockout U2OS cells reached 50% confluency in 6-well plates, they were transfected with 0.5 µg of a PiggyBac-compatible expression vector and 0.2 µg 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 1 µg/ml puromycin (Sigma). Cells were cultured in selection media for several weeks until the polyclonal pool of integrants had reached confluency. Cell lines were immediately cryopreserved and cultured in selection medium containing 60 µg/ml cumate (System Biosciences) in preparation for analysis. Cell lines integrated with the lysosomal degradation assay reporter (pLGW733) were treated as above with the exception that they were selected for in 200 µg/ml G418 as opposed to puromycin and were not cultured in the presence of cumate.

2.2.2.3 Cell Assays

2.2.2.3.1 Lectin Stains

Wild-type and various FAM114A and GOLPH3 family knockout 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 consisting of 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 8 fluorescein-labelled lectins (see Table 2.1.3.3) and fixable

viability dye eFluor 780 (1:1000, Thermo Fischer Scientific) diluted in FACS buffer on ice in darkness for 30 minutes. 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). Data was analysed and histogram plots generated using FlowJo V10 in which singlets were gated for while dead cells were excluded from analysis. Compensation was conducted using single colour control samples.

2.2.2.3.2 *In Vivo* 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 Alexa Fluor 647-conjugated anti-GFP antibody and fixable viability dye eFluor 780 diluted in FACS buffer on ice in darkness for 30 minutes. Cells were subsequently washed, fixed and analysed as described for lectin stains (see 2.2.2.3.1).

2.2.2.3.2 MitoID Proximity-Dependent Labelling Assay

Doxycycline-inducible stable Flp-In T-REx HEK293 cell lines expressing mitochondrial-relocated golgin-BirA* fusion proteins were induced with 1 µg/ml doxycycline (Sigma) in culture medium once they reached approximately 80% confluency. 24 hours after transfection, cells were treated with 0.5 µM nocodazole (Sigma), 50 µM biotin (Sigma) and 1 µg/ml doxycycline in culture medium for a further 9 hours. Cells were harvested and lysed in preparation for a streptavidin pulldown (see 2.2.6.4).

2.2.2.3.3 Lysosomal Degradation Assay and Rescues

For genetic and pharmacological rescue experiments, wild-type and *ΔΔGOLPH3, GOLPH3L* U2OS cells were seeded at a density of $2x10^4$ cells/cm^2 in 6-well plates in culture medium in a humidified incubator at 37°C with 5% CO_2. 24 hours after seeding, cells were transfected with wild-type or RR14,15AA mutant GOLPH3 expression constructs as described previously (see 2.2.2.1). Alternatively, cells were cultured in the presence of Bafilomycin (160 nM, Sigma) or MG-

132 (10 μM, Sigma) for 24 hours or 4 hours respectively. Cells were trypsinated and harvested for western blot analysis.

For the lysosomal degradation assay, a reporter consisting of the CTS domain of GALNT2 fused to GFP, a P2A-T2A sequence and mCherry was stably integrated into wild-type or *ΔΔGOLPH3, GOLPH3L* U2OS cell lines (see 2.2.2.2.3). Polyclonal cell lines were selected for in 200 μg/ml G418 in culture medium. Once cells reached 80-90% confluency in T-75 flasks, 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 15 ml falcon tube and centrifuged at 300 x g for 5 mins. Cells were resuspended in ice cold FACS buffer prior to immediate flow cytometry analysis as described lectin stains (see 2.2.2.3.1).

2.2.3 Insect Cell Biology

2.2.3.1 Cell Culture, Cryopreservation and Transfection

D.Mel-2 cells were maintained in Schneider's Drosophila Medium (Thermo Fischer Scientific) with 10% FBS and pen-strep at 24°C. Cells were subcultured at a ratio of 1:10 every 3-4 days in which loosely attached cells were detached through tapping of the flask and the cell suspension diluted in fresh medium and transferred to a fresh flask. Insect cells were cryopreserved as described for mammalian cells. Large scale D.Mel-2 cultures were prepared by diluting cells to a density of 10^6 cells/ml in Insect Xpress (Lonza) culture medium in Erlenmeyer flasks. Cells were incubated at 25°C with shaking at 140 rpm and subcultured by dilution at a ratio of 1:10 every 3-4 days. Sf9 cells were cultured and prepared for order by the MRC LMB Baculovirus Facility.

2.2.3.2 Baculoviral Expression System

Sf9 cells were seeded at a density of 10^6 cells/cm^2 in 6-well plates in Insect Xpress culture medium at 27°C and allowed to adhere for at least 10 minutes. Cells were transfected with 2 μg of bacmid DNA using Fugene HD transfection reagent (Promega) according to the manufacturer's protocol. Cells were incubated for 3-5 days to allow for sufficient propagation of the virus. The medium containing the virus was aspirated from the wells and used to inoculate a 50 ml culture of cells at a density of $2x10^6$ cells/ml in 250 ml Erlenmeyer flasks. Cells were cultured at 27°C with shaking at 140 rpm for 3-5 days. Cells were pelleted by centrifugation at 2500 x g for 10 minutes and the pellet, containing cells expressing the recombinant protein, was stored on ice or snap frozen in liquid nitrogen prior to protein purification. Alternatively, the supernatant containing the virus was used to inoculate larger cultures or to enhance the viral titre. The supernatant was preserved in the presence of 2% FBS at 4°C in darkness for medium term storage.

2.2.4 Immunofluorescence Microscopy

Cells were seeded onto multispot coated microscope slides (Hendley-Essex) in culture medium in a humidified incubator at 37°C with 5% CO_2. Cells were cultured for 8-20 hours to allow sufficient adherence to the slides. Where applicable, cells were subsequently cultured in medium containing 0.5 µM nocodazole for 4-6 hours. Slides were moved to room temperature and were washed twice in PBS and fixed in 4% PFA in PBS for 20 minutes. Cells were washed a further two times in PBS prior to permeabilisation in 10% Triton X-100 (Thermo Fischer Scientific) in PBS for 10 minutes. Cells were incubated in blocking buffer consisting of 20% FBS, 1% Tween-20 (Sigma) in PBS for 1 hour. Blocking buffer was aspirated and cells were incubated in primary antibody diluted in blocking buffer for 1 hour (see Table 2.1.3.1). The primary antibody was aspirated and cells were washed twice with PBS, incubated for 10 minutes in blocking buffer and washed a further two times with PBS. Cells were incubated in secondary antibody diluted in blocking buffer for 1 hour (see Table 2.1.3.2) and subsequently washed as described for the primary antibody. PBS was aspirated and cells were mounted in VECTASHIELD mounting media (Vector Laboratories) prior to application of the coverslip. Coverslips were sealed using nail vanish prior to imaging using the Leica TCS SP8 confocal microscope. Images were processed using Adobe Photoshop CC 2017.

2.2.5 Biochemistry

2.2.5.1 Protein Sample Preparation

Whole adult flies and pelleted mammalian, bacterial and insect 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 the cOmplete EDTA-free protease inhibitor cocktail (Roche). Large scale cultures were sonicated on ice for 1 minute; 10 seconds on, 10 seconds off (bacteria cells) or for only 10 seconds (mammalian and insect cells) at 45% amplitude using a Sonic Vibra-Cell lance sonicator. Smaller scale cultures were sonicated using the Misonix 300 water sonicator for 1 minute (mammalian and insect cells) or 3 minutes (bacteria cells); 10 seconds on, 10 seconds off at amplitude 5.0. Cells were placed on fresh ice for at least 5 minutes and incubated with agitation at 4°C for a further 10 minutes. Whole adult flies were homogenised in lysis buffer using a motorised pestle. lysates were clarified by centrifugation at 16,100 (small scale) or 32,000 (large scale) x g for 10 minutes at 4°C. Where appropriate, protein content was quantified using the Pierce BCA Protein Assay Kit (Thermo Fischer Scientific) and Infinite F200 plate reader (Tecan) and subsequently normalised across treatments. Protein samples were kept on ice prior to downstream purification or were resolved by SDS PAGE.

2.2.5.2 Resolution of Protein Samples

Protein samples were incubated at 90°C for 3 minutes in 1x NuPage LDS sample buffer (Thermo Fischer Scientific) with 10% β-mercaptoethanol (Sigma) and loaded into Novex 4-20% Tris-Glycine Mini Gels (Thermo Fischer Scientific). Gels were resolved using the XCell SureLock Mini-Cell electrophoresis system (Thermo Fischer Scientific) for 1 hour at a constant voltage of 175 V in tris-glycine SDS running buffer (see Table 2.2.5.2). 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 western blot in which protein was transferred onto a 0.45 µm nitrocellulose membrane (GE Life Sciences) using the Mini Trans-Blot Cell (Bio-Rad) in transfer buffer (see Table 2.2.5.2), 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 (see Table 2.1.3.1) diluted in 3% milk in PBS for 2 hours at room temperature or overnight at 4°C. Blots were washed four times for 5 minutes in PBST and incubated with the secondary antibody (see Table 2.1.3.2) diluted in 3% milk in PBS for 1 hour at room temperature. Blots were washed four times for 5 minutes in PBST. ECL western blotting detection reagent (GE Life Sciences) was applied to blots for 3 minutes prior to exposure to X-ray films (Photon Imaging Systems). The films were developed using a JP-33 film processer (JPI Healthcare Solutions). Alternatively, western blots were probed with fluorescein-conjugated lectins (see Table 2.1.3.3) as described previously (Sato, 2014) and fluorescent blots were visualised using the Amersham Typhoon Biomolecular Imager (GE Life Sciences). The constituents of the buffers used for SDS PAGE and western blot transfer are given in table 2.2.5.2.

Table 2.2.5.2

Buffer	Constituents
Tris-glycine running buffer	0.3% w/v Tris base 1.44% w/v Glycine 20% v/v SDS In H_2O
Transfer buffer	0.015% w/v Tris base 0.072% w/v Glycine 20% v/v Methanol In H_2O

2.2.5.3 GST Affinity Chromatography

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 lysate (see 2.2.5.1) from strains expressing GST-tagged fusion proteins was 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. Where downstream samples were required for mass spectrometry analysis, mammalian and insect cell lysates were preincubated on glutathione sepharose beads at 4°C for 30 minutes with agitation to preclear proteins prone to non-specific binding. Mammalian and insect cell lysates were mixed with the beads loaded with GST-tagged proteins and incubated for 1 hour with agitation 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. Bait proteins were eluted by boiling in 2x LDS with 10% β-mercaptoethanol.

2.2.5.4 MitoID Streptavidin Pulldown

The list of buffers used in the MitoID streptavidin pulldown is given in Table 2.2.5.4. Dynabeads One Streptavidin T1 beads (Thermo Fischer Scientific) were washed once in lysis buffer with the aid of a DynaMag-2 magnetic stand (Thermo Fischer Scientific). Cell lysates containing biotinylated proteins were loaded onto washed beads and incubated overnight with agitation at 4°C. The beads were washed twice in wash buffer 1 for 8 minutes, 3 times in wash buffer 2 for 8 minutes and 3 times in wash buffer 3 for 8 minutes. Proteins were eluted by boiling at 98°C in 1x LDS, 10% β-mercaptoethanol and 6 mM biotin for 5 minutes.

Table 2.2.5.4

Buffer Name	Constituents
Lysis buffer	50 mM Tris HCl pH 7.4 150 mM NaCl 1 mM EDTA 0.5% Triton X-100 1 mM PMSF 1x cOmplete EDTA-free protease inhibitor
Wash buffer 1	2% SDS 1x cOmplete EDTA-free protease inhibitor
Wash buffer 2	50 mM Tris HCl pH 7.4 500 mM NaCl 1 mM EDTA 1% Triton X-100 0.1% deoxycholate 1x cOmplete EDTA-free protease inhibitor
Wash buffer 3	50 mM Tris HCl pH 7.4 50 mM NaCl 1x cOmplete EDTA-free protease inhibitor

2.2.5.5 Tandem Lectin Affinity Chromatography

1 ml of clarified lysate from 50 whole adult flies per genotype was mixed with 20 µg of biotinylated-jacalin and incubated for 1 hour at 4°C with agitation. 100 µl of Dynabeads One Streptavidin T1 beads were washed once in lysis buffer with the aid of a DynaMag-2 magnetic stand. Lysate-lectin mixtures were loaded onto washed beads and incubated for 30 minutes with agitation at 4°C. The beads were isolated by magnetism and the flow through was retained and subject to a second round of lectin affinity in which it was mixed with 20 µg of biotinylated-VVA. Lysate-lectin mixtures were incubated and subsequently loaded onto streptavidin beads as described for jacalin. For both lectins, magnetism was utilised to wash beads four times in lysis buffer and proteins that specifically associated with the lectins were eluted by resuspension in lectin elution buffers containing high concentrations of specific competitive sugars (see Table 2.2.5.5). Beads were incubated for 5 minutes at room temperature with agitation in elution buffer and magnetism was utilised to isolate the eluate. Competitive sugar elutions were subject to TCA/acetone reprecipitation and were resolubilised in 1x LDS with 10% β-mercaptoethanol.

Table 2.2.5.5

Buffer Name	**Constituents**
Lysis buffer	50 mM Tris HCl pH 7.4 150 mM NaCl 1 mM EDTA 0.5% Triton X-100 1 mM PMSF 1x cOmplete EDTA-free protease inhibitor
Jacalin Elution Buffer	50 mM Tris HCl pH 7.4 500 mM NaCl 0.5% Triton X-100 800 nM galactose
VVA Elution Buffer	50 mM Tris HCl pH 7.4 500 mM NaCl 0.5% Triton X-100 200 mM *N*-acetylgalactosamine

2.2.6 Mass Spectrometry

Protein samples generated from the MitoID assay, GST affinity chromatography and tandem lectin affinity chromatography were resolved by SDS PAGE and gels were stained as described in 2.2.5.2. Protein bands were excised for subsequent trypsin digestion and analysis by Nanoflow reverse-phase liquid chromatography-mass spectrometry using the Velos Orbitrap mass spectrometer (Thermo Fischer Scientific) as described previously (Gillingham et al., 2019). For spectral count analysis of the results of the MitoID assay and tandem lectin affinity chromatography, Mascot (Matrix Science) was used to search for peptides against the UniProt

human proteome and further filtered using Scaffold (Proteome Software Inc). MitoID spectral counts were compared using D-score analysis from the open-source ComPASS platform (Sowa et al., 2009). Total spectral counts from tandem lectin affinity chromatography were used to calculative the relative abundance values of each protein for both lectins. Total spectral counts from samples of two independent *ΔCG9590* lines were averaged and divided by the spectral counts from a sample from wild-type flies. Mass spectrometry data generated from GST affinity chromatography was analysed using MaxQuant and peptides were searched for against the UniProt human proteome using Andromeda (Cox et al., 2011; Cox and Mann, 2008). The Perseus platform was used to filter samples and to convert protein LFQ intensities to logarithmic values (Tyanova et al., 2016). Missing values were subsequently imputed using the default settings and statistical comparisons of treatments were made using Student's t-tests. Volcano plots were also generated using the Perseus platform. Trypsin digestions, mass spectrometry and proteome searches were conducted by Farida Begum, Sew-Yeu Peak-Chew and Mark Skehel in the MRC LMB Mass Spectrometry Core Facility.

2.2.7 *Drosophila Melanogaster* Genetics

To remove the entire coding region of the fly ortholog of FAM114A, *CG9590*, embryos expressing Cas9 in the germline were injected with either gRNA plasmids pCFD3_1 and pCFD_3 or gRNA plasmids pCFD3_2 and pCFD3_4 both sets targeting sequences upstream and downstream of *CG9590*. After injections embryos were kept at 25°C to develop. The adult G0 flies were crossed to balancer flies to establish stable stocks. In total, 200 single crosses were set up for genotyping. The DNA was extracted and diagnostic PCRs were used to recover deletions of the desired region and size. To validate the recovered deletions and determine the precise breakpoints PCR products were subsequently sequenced. In summary, three stable fly stocks were recovered, FAM114A_Δ31, FAM114A_Δ43 and FAM114A_Δ76, all of which completely remove *CG9590*. Only two lines, FAM114A_Δ31 and FAM114A_Δ76, were used in the present study. The generation and maintenance of fly stocks and the writing of the associated methods was carried out by Nadine Muschalik.

2.2.8 *In Silico Analyses*

The amphipathic helical character of protein sequences was calculated, and associated helical plots generated, using HeliQuest (Gautier et al., 2008). Multiple sequence alignments were conducted using MUSCLE (Edgar, 2004) or Clustal Omega (Sievers and Higgins, 2014) analysis and alignment plots generated using SnapGene or BoxShade using to the default settings. Secondary structure and intrinsic protein disorder was predicted using PSIPRED V4.0 and DISOPRED 3 respectively, cartoons were generated using the PSIPRED Workbench (Buchan and Jones, 2019; Jones and Cozzetto, 2015).

3. Identification of FAM114A proteins as novel early-Golgi COPI vesicle-residents

3.1 Introduction

One of the key aims of this thesis is to understand how different Golgi-resident COPI cargo is differentially distributed across the Golgi stack. To tackle this question, it is necessary to better understand the diversity of intra-Golgi COPI vesicles. In order to address this, a proteomic screen was employed to identify novel intra-Golgi vesicle-resident factors.

3.2 A MitoID screen to identify novel intra-Golgi vesicle-resident proteins reveals FAM114A2

The large coiled-coil intra-Golgi golgin tethers, GMAP-210 and golgin-84 have been previously shown to be important for tethering intra-Golgi COPI-dependent vesicles (Malsam et al., 2005; Wong and Munro, 2014). More specifically, the extreme N-terminus of GMAP-210 and golgin-84 have been demonstrated to be both necessary and sufficient for intra-Golgi vesicle capture (Figure 3.1A; Wong et al., 2017). Furthermore, a conserved tryptophan residue when targeted by missense mutation had been shown to ablate vesicular capture for golgin-84 and cause GMAP-210 to lose the ability to capture giantin- and GALNT2-containing vesicles but not golgin-84-containing vesicles (Wong et al., 2017).

In order to identify novel intra-Golgi vesicle-resident factors, I applied the unbiased MitoID proximity-dependent labelling assay to the intra-Golgi golgin tethers GMAP-210 and golgin-84 (Gillingham et al., 2019; Shin et al., 2017). The basis of the screen was to ectopically relocate intra-Golgi vesicles to the mitochondria and label the vesicle-resident proteins for subsequent purification and identification by mass spectrometry (Figure 3.1B). The wild-type N-termini of GMAP-210, golgin-84 and analogous tryptophan missense mutants were fused to BirA*-GCC185-MAO (Figure 3.1A). GCC185, a large coiled-coil golgin previously shown not to exhibit ectopic vesicle capture in the cell lines used in the present study (Wong and Munro, 2014), was used as an inert platform to enable long distance vesicle tethering by the N-terminal fragments. Fusion to the promiscuous biotin ligase BirA* enabled proximity-dependent biotinylation of vesicle-residents (Roux et al., 2012). Moreover, cells were treated with nocodazole to depolymerise microtubules and scatter the Golgi into several ministacks to improve the proximity of intra-Golgi carriers to the mitochondria in order to enhance the efficiency of capture (Wong and Munro, 2014). Fusion of the C-terminal transmembrane domain of monoamine oxidase A (MAO) permitted

ectopic localisation of the golgin-fusions and associated tethered intra-Golgi transport vesicles to the mitochondria, away from the Golgi. This provided a clear basis for assaying capture and dissecting the proximal milieu of biotinylated proteins.

The Flp-In T-REx system was utilised to generate HEK293 cells capable of inducible expression of the golgin-BirA* chimeras to serve as a platform upon which to conduct the proteomic screen for novel vesicle-residents (Figure 3.1B). Expression was induced in cells for 24 hours prior to a 9-hour nocodazole treatment and biotin pulse. Cells were lysed and biotinylated proteins were purified by streptavidin affinity chromatography and identified by mass spectrometry. Using total spectral counts for each protein, D-scores were calculated which incorporates the uniqueness and abundance of each protein within the data set (Sowa et al., 2009). As a part of the chimeric bait itself, GCC185 displayed one of the highest D-scores across many of the samples (Figure 3.1C). Several known COPI vesicle cargo proteins were enriched in the wild-type golgin-84 and/or GMAP-210 N-terminal chimera samples including endogenous golgin-84, giantin and ERGIC-53. Another closely related intra-Golgi golgin, TMF, was enriched in the wild-type golgin-84 N-terminal chimera sample. Of the top-scoring hit proteins which presented as a potential novel vesicle-resident protein, protein of unknown function FAM114A2 was enriched in both repeats for the wild-type GMAP-210 N-terminal chimera and in one repeat for the wild-type golgin-84 N-terminal chimera sample. The orthologous *Drosophila* FAM114A protein was initially identified as a novel Rab2 effector in a proteomic screen using S2 cells which was subsequently supported by findings in a Rab MitoID screen in mammalian cells (Gillingham et al., 2014; Gillingham et al., 2019). The orthologous *Drosophila* protein and two orthologous human proteins, FAM114A1 and FAM114A2, were also shown to colocalise with cis-Golgi markers (Gillingham et al., 2014).

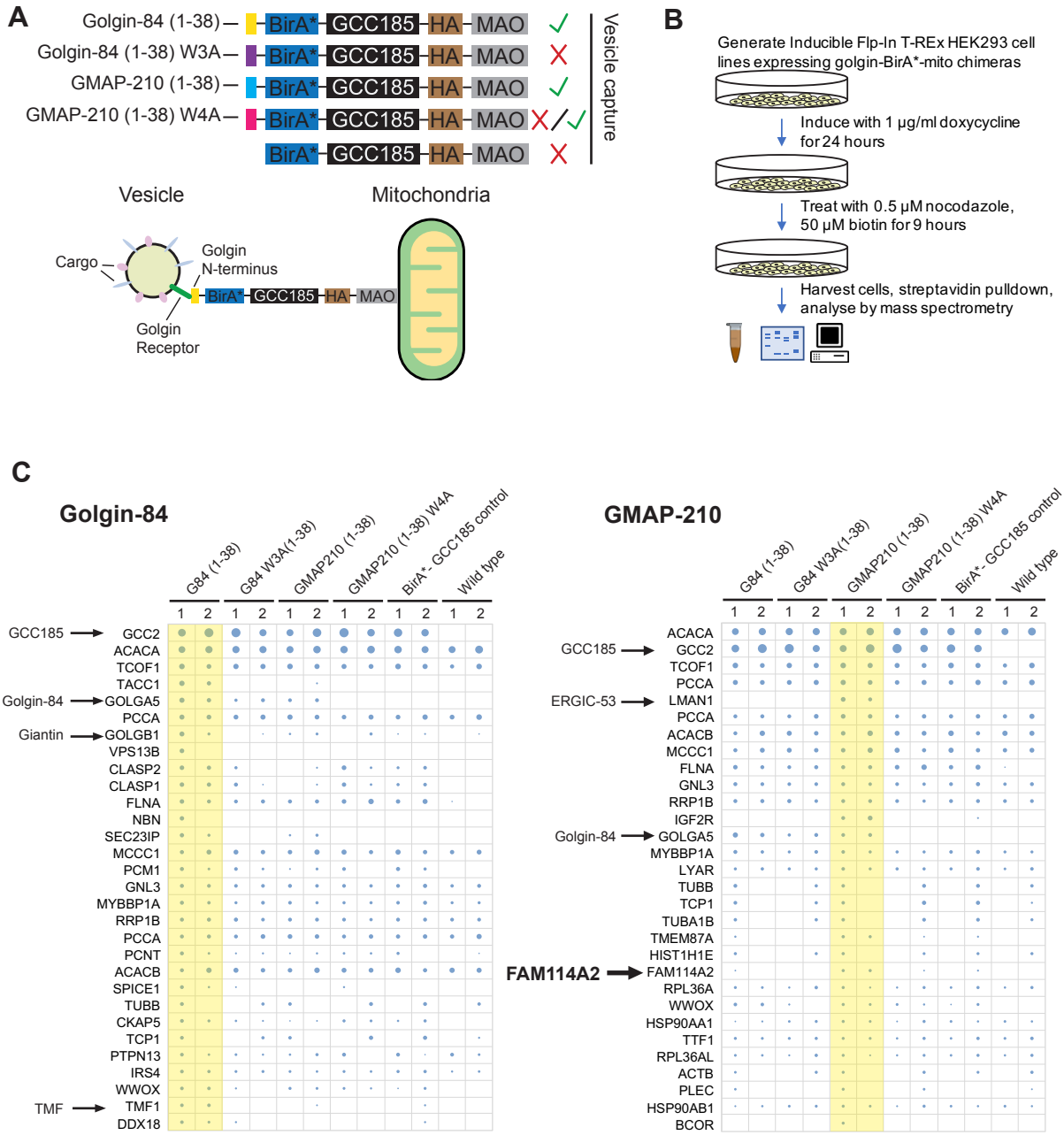


Figure 3.1 A MitoID screen to identify novel intra-Golgi vesicle-resident proteins reveals FAM114A2. (**A**) A cartoon diagram outlining vesicle capture capabilities of golgin-BirA* chimeras. GMAP-210 W4A mutants lose the ability to capture giantin- and GALNT2-containing vesicles but not golgin-84-containing vesicles. (**B**) A schematic representing the strategy of the MitoID screen. The 9 hour nocodazole treatment and biotin pulse were required to facilitate ectopic vesicle capture through Golgi fragmentation and to promote proximity-dependent biotinylation respectively. (**C**) D-score analysis of mass spectrometry MitoID data. The size of the bubble corresponds to the D-Score assigned to each protein (corresponding gene name, left axis) for each golgin-BirA* GCC185-HA-MAO bait (top axis, upper) and for each repeat (top axis, lower). Wild-type corresponds to inducible Flp-In T-REx cells integrated with an empty pcDNA5/FRT/TO vector expressing no bait. Presented is same data set showing the top 30 D-scores aligned for the baits golgin-84 (1-38) and GMAP-210 (1-38) -BirA*-GCC185-HA-MAO on the left and the right respectively.

3.3 FAM114A proteins are likely to specifically associate with membranes of the early-Golgi

While the early-Golgi localisation of FAM114A proteins has been demonstrated, their function and role as novel peripheral intra-Golgi vesicle proteins has not been previously explored. In order to explore the possibility that FAM114A proteins specifically reside on COPI vesicles recycling between early-Golgi cisternae, GFP-tagged FAM114A proteins were overexpressed in HeLa cells and counterstained with the COPI cargo protein and ERGIC marker ERGIC-53 (Figure 3.2A). GFP-tagged FAM114A2, but not FAM114A1, showed considerable colocalisation with ERGIC-53 suggesting FAM114A2 may reside on early-Golgi membranes and/or early intra-Golgi vesicles. Previous studies have shown overexpressed GFP-tagged FAM114A1 colocalises with cis-Golgi marker GM130 (Gillingham et al., 2014). As the mitochondrial relocation assay has been successfully leveraged to explore the capacity of intra-Golgi golgins to interact with vesicles (Wong and Munro, 2014; Wong et al., 2017), the same technique was applied to the FAM114A proteins. Both FAM114A1 and FAM114A2, when ectopically-relocated to the mitochondria in HeLa cells, caused the rerouting of a number of COPI cargo proteins including giantin, GALNT2 and golgin-84 (Figure 3.2B). The relocation of FAM114A proteins triggered a striking change in mitochondrial distribution. In the BioID2-HA-MAO negative control sample, mitochondria exhibited a typical tubular pattern whereas mitochondria formed large clumps (approximately 10-20 μm wide) with FAM114A mitochondrial relocation; a sign which is often indicative of vesicular capture (Wong and Munro, 2014). Furthermore, the mitochondrial-relocation of FAM114A proteins caused considerable fragmentation of the rest of the Golgi. While the trans-Golgi marker golgin-245 displayed a classic perinuclear ribbon morphology in the BioID2-HA-MAO negative control, in FAM114A-mitochondrial-relocated cells it displayed a fragmented pattern which was decorated around the mitochondrial clump.

Giantin and golgin-84 are both COPI vesicle cargo proteins which are also localised at the cis-Golgi rims (Diao et al., 2003; Sönnichsen et al., 1998; Tie et al., 2018). Moreover, GALNT2 is responsible for the initial modification in mucin-type O-linked glycosylation and therefore would also be expected to reside in early-Golgi compartments (Wandall et al., 1997). It is therefore likely that these cargo proteins are associated with COPI vesicles which cycle between early cisternae. In contrast, TGN46 is suggested to partially reside in COPI vesicles of the late-Golgi (Adolf et al., 2019; Wong and Munro, 2014) while GOLPH3 is a COPI cargo adaptor for vesicles budding from the trans-Golgi (Wood et al., 2012). Unlike early-Golgi cargo, TGN46 did not display direct relocation when FAM114A1 was

relocated to the mitochondria but rather colocalised with golgin-245 around the edge of the mitochondrial clump (penultimate row, Figure 3.2B). GOLPH3 was also not rerouted to the mitochondrial clump in the presence of mitochondrial-relocated FAM114A2 but displayed marked depletion when compared to neighbouring cells not expressing the FAM114A2 fusion (last row, Figure 3.2B). This suggests that FAM114A proteins selectively relocate cargo associated with COPI vesicles of the early-Golgi as opposed to the late-Golgi. Of the markers tested, no distinction was observed between the COPI cargo FAM114A1 and FAM114A2 could reroute.

COPI cargo must first be synthesised at the ER and traffic to the Golgi via COPII-dependent vesicular transport before it is recycled in retrograde COPI vesicles. Therefore, in order to ascertain whether the mitochondrial-relocated FAM114A proteins are capturing anterograde COPII-dependent vesicles, an assay was devised combining the retention using selective hooks (RUSH) system (Boncompain et al., 2012) and the mitochondrial-relocation assay (Figure 3.2C). The basis of the assay was to add biotin to trigger the synchronous release of an EGFP-tagged GALNT2 reporter from the ER in COPII vesicles and to test if these vesicles could be captured by mitochondrial-relocated FAM114A proteins. In cells expressing just the GALNT2 reporter, 30 minutes after the addition of biotin the reporter had completely reached the Golgi (micrographs, Figure 3.2C). In neighbouring cells which were expressing both the reporter and mitochondrial-relocated FAM114A2, the reporter failed to exit the ER after 30 minutes. It remains to be seen if this phenomenon is indicative of a wide-spread hinderance of ER export or is specific to the reporter. Furthermore, it was not possible to directly ascertain if mitochondrial-relocated FAM114A proteins were capturing anterograde COPII vesicles. Regardless, the attenuation of ER export and the fragmentation of the Golgi ribbon together demonstrate that the mitochondrial-relocation of FAM114A proteins causes significant impairments to the early secretory pathway.

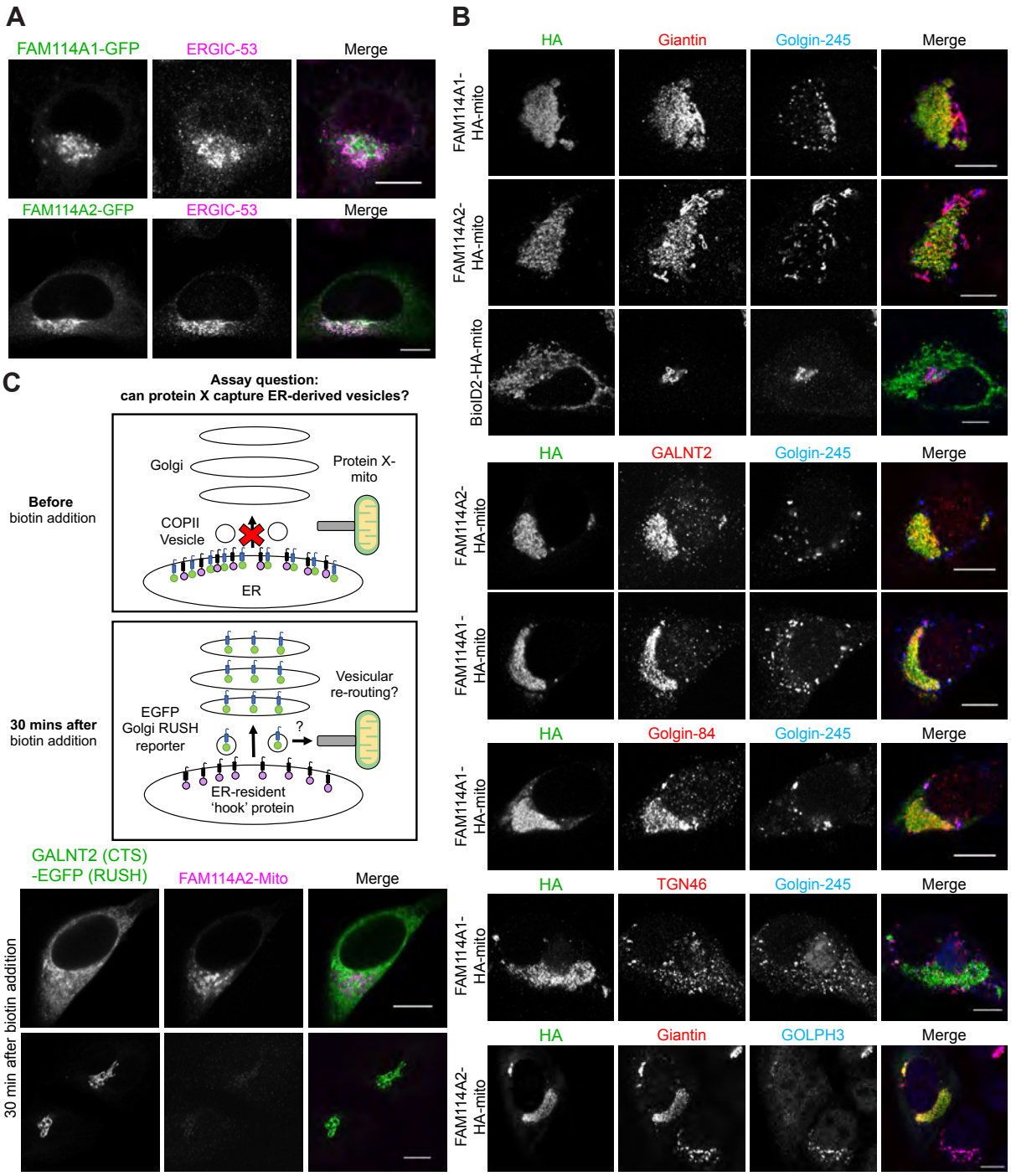


Figure 3.2 Ectopic mitochondrial relocation of ERGIC/early-Golgi resident FAM114A proteins causes the recruitment of intra-Golgi cargo and the disruption of the early secretory pathway. HeLa cells expressing GFP-tagged FAM114A proteins (**A**) or mitochondrial-relocated HA-tagged FAM114A proteins (**B**). Cells were counterstained for different intra-Golgi cargo proteins to assay the range of membrane capture capability of FAM114A proteins. In most instances, cells were counter-stained with the trans-Golgi marker golgin-245 to visualise the remainder of the Golgi. In the final row, cells were simultaneously stained for giantin and GOLPH3 as an early and late intra-Golgi vesicular marker respectively. BioID2-HA-mito was included as a negative control. (**C**) A schematic describing an assay to investigate the capability of a protein to capture ER-derived vesicles. The RUSH system was used to track EGFP-tagged reporters immediately after release from the ER on the addition of biotin. The intention was to visualise the rerouting of the reporter to the mitochondria in the presence of a mitochondrial-relocated protein of interest. In the presence of mitochondrial-relocated FAM114A2, ER export of a GALNT2 reporter was inhibited. Micrographs were collected from the same well, 30 minutes after the addition of 40 μM biotin. Scale bars correspond to 10 μm.

3.4 FAM114A proteins contain a predicted ALPS which is necessary for recruitment to the membrane

FAM114A proteins are highly-conserved across several eukaryotic kingdoms with the notable exception of fungi (Figure 3.3A). Both mammalian paralogs share a domain of unknown function 719 (DUF719) and a small predicted region of coiled-coil. At the N-terminus of the DUF719 is an evolutionary-conserved block of residues characterised by a repeat of a tryptophan and glycine residue (WG repeat). Analysis of the amphipathic nature of the WG repeat and neighbouring residues predicts a high propensity to form an amphipathic helix with the tryptophan residues serving as key residues on the hydrophobic face (right panels, Figure 3.3A). The resulting helix resembled an amphipathic lipid packing sensor (ALPS) motif which have been shown to absorb onto highly-curved membranes (Bigay et al., 2005). They are characterised by their polar uncharged residues (serines and theonines) on their hydrophilic face and bulky hydrophobic residues (tryptophans, leucines and phenylalanines) on the opposite face (Antonny, 2011). In order to interrogate the potential ALPS motifs, the three tryptophans of the hydrophobic face in the WG repeat of FAM114A1 and FAM114A2 were mutated to alanines which is predicted to ablate the amphipathic character of the motif ('WG' mutants). When the GFP-tagged 'WG' mutant form of FAM114A2 was overexpressed in HeLa cells, it no longer colocalised with the Golgi marker giantin (Figure 3.3B). Moreover, both 'WG' mutant FAM114A proteins lost the ability to reroute the intra-Golgi cargo protein giantin to the mitochondria (Figure 3.3C). Despite a loss of capture capability, the mitochondrial-relocated FAM114A1 'WG' mutant was still able to induce mitochondrial clumping and Golgi fragmentation. The Golgi fragments appeared to intercalate amongst the mitochondrial clump but showed no direct colocalisation. In contrast, the mitochondrial-relocated FAM114A2 'WG' mutant was no longer able to induce mitochondrial clumping and the Golgi displayed a normal perinuclear ribbon morphology. To ask whether the observed effects of the mutations were due to disruption of the predicted ALPS motif as opposed to the loss of a sequence specific interaction (e.g. protein-protein interaction), a chimera was generated in which the ALPS motif of NUP133, a functionally unrelated nucleoporin protein, was substituted for that of FAM114A1 (Figure 3.3D). Both mitochondrial-relocated wild-type FAM114A1 and the NUP133 chimera were able to reroute golgin-84 to the mitochondria, induce mitochondrial clumping and cause the rest of the Golgi to fragment. This provides convincing evidence that FAM114A proteins have a conserved ALPS motif which is necessary for interactions with what are presumably highly-curved Golgi membranes.

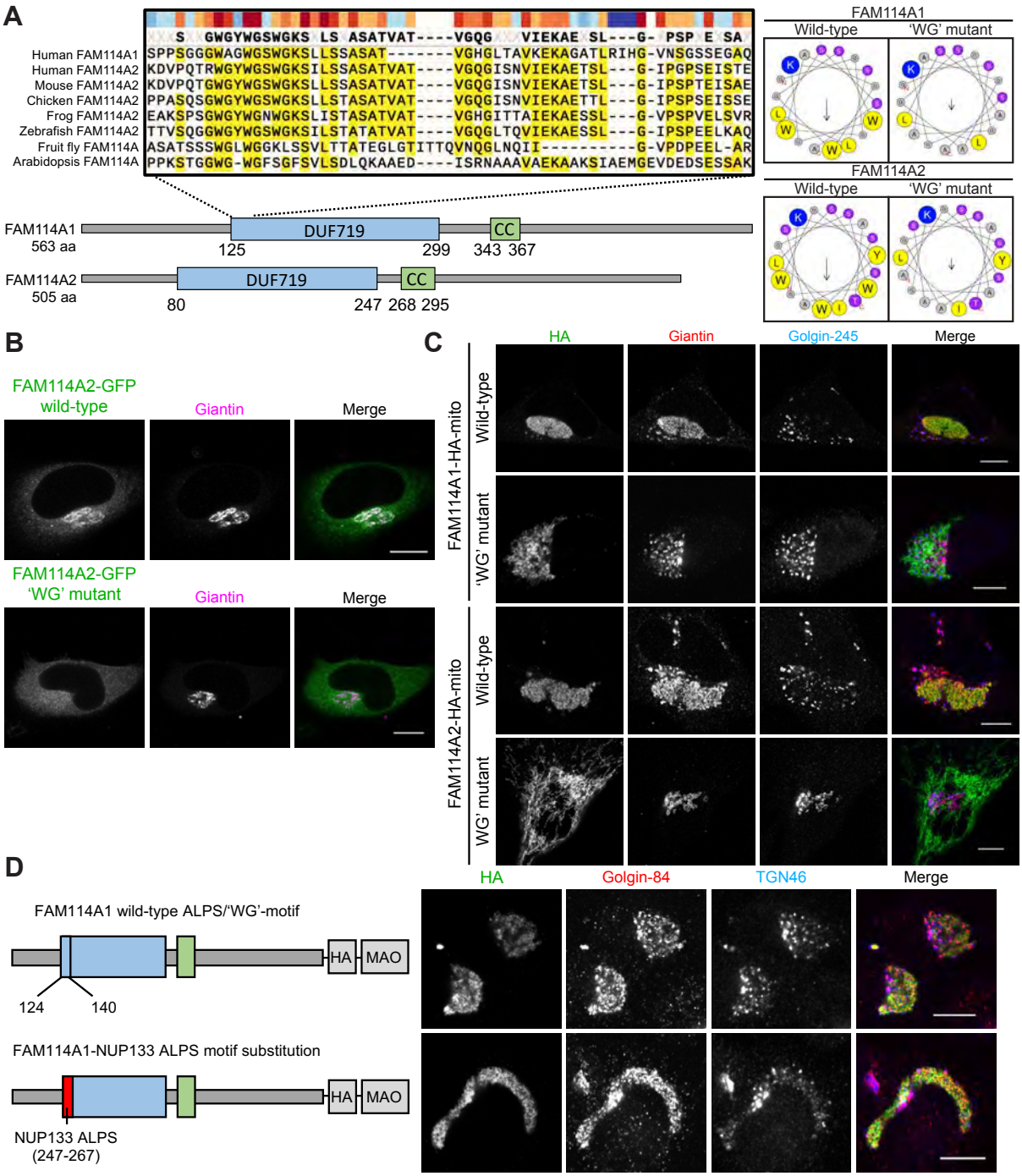


Figure 3.3 FAM114A proteins contain a highly conserved 'WG' repeat at the start of the domain of unknown function (DUF719) which resembles an ALPS motif and is necessary for vesicular capture and Golgi-targeting. (**A**) The domains of FAM114A proteins. MUSCLE protein sequence alignment of the N-terminus of the DUF719 of FAM114A homologues reveals a conserved 'WG' repeat and high sequence conservation in neighbouring residues. HeliQuest analysis of these residues predicts the formation of an amphipathic lipid packing sensor (ALPS) motif (FAM114A1 124-140, FAM114A2 82-99). The tryptophan residues of the 'WG' repeat and of the predicted hydrophobic face were subject to alanine mutagenesis. CC, a predicted region of coiled-coil. (**B**) HeLa cells expressing wild-type and 'WG' mutant GFP-tagged FAM114A2 and stained for giantin (Golgi). (**C**) HeLa cells expressing HA-tagged mitochondrial-relocated wild-type or 'WG' mutant FAM114A proteins. Cells were stained for giantin (intra-Golgi cargo) and golgin-245 (trans-Golgi). (**D**) The ALPS motif of FAM114A1 was substituted for that of the functionally unrelated protein NUP133. MAO, a mitochondrial-targeting sequence. The ability of the chimera to relocate intra-Golgi cargo was compared to that of the wild-type protein. Cells were stained for golgin-84 (intra-Golgi cargo) and TGN46 (late-Golgi). Scale bars correspond to 10 μm.

3.5 Regions outside of the ALPS motif of FAM114A proteins confer specificity to membrane-binding

While the ALPS motif of FAM114A proteins is necessary in order to reroute COPI cargo to the mitochondria, it does not explain how FAM114A proteins bind specifically to membranes associated with the early-Golgi. Also, the ALPS motif makes up a small part of the highly conserved DUF719 and is not clear how it functions in the context of the whole domain. In order to investigate this further, the DUF719 of FAM114A1 (residues 125-299) and FAM114A2 (residues 80-247) were tested for their ability to relocate giantin to the mitochondria (Figure 3.4A). The DUF719 of FAM114A1 was able to relocate giantin to the mitochondria and cause the Golgi to fragment but with a lesser efficiency than the full-length protein (Figure 3.2, 3.3, 3.4A). In contrast, the mitochondrially-relocated DUF719 of FAM114A2 was insufficient to relocate giantin or fragment the Golgi (Figure 3.4A). When expressed at high levels, the DUF719 of FAM114A2 could induce mitochondrial clumping but unlike the mitochondrial clumps observed for the full-length protein, the clump appeared to be intercalated with large vesicular structures. When GFP-tagged and overexpressed in HeLa cells, the DUF719 of FAM114A1 colocalised with giantin and displayed an ER and/or cytoplasmic pattern (Figure 3.4B), a localisation profile which is broadly comparable to the full-length protein. This was again in contrast with the DUF719 of FAM114A2, which when GFP-tagged and overexpressed, displayed only minimal colocalisation of with giantin. These findings begin to suggest that while the ALPS motif can interact directly with highly-curved membranes, other parts of the FAM114A proteins may be responsible for increasing the efficiency and specificity by which they interact with membranes. To test this, a shorter region of FAM114A2 which contains the ALPS motif (residues 73-133) was GFP-tagged and overexpressed to see if the specificity of membrane binding could be removed (Figure 3.4C). The truncation mutant colocalised with giantin but also appeared to localise to other intracellular compartments (see inset, Figure 3.4C) and its overexpression triggered the fragmentation of the Golgi. To confirm this broad-localisation profile was dependent on the ALPS motif, a WG mutant form of the truncation was also GFP-tagged and overexpressed. This mutant no longer localised to the same intracellular compartments but instead displayed a diffuse cytosolic and nuclear pattern. In summary, this suggests that the predicted ALPS motif is necessary for membrane association but other domains are important to contribute specificity to this membrane interaction.

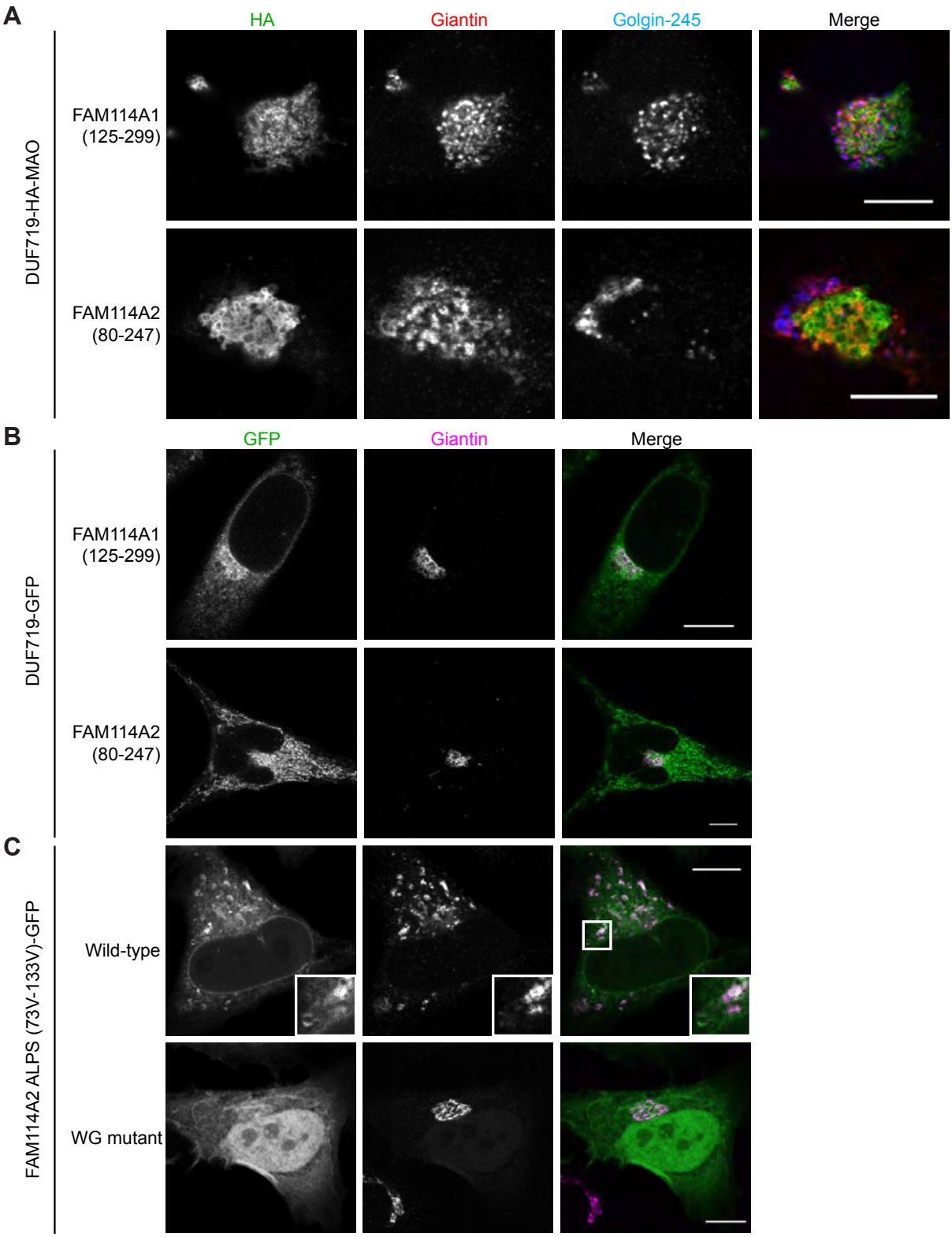


Figure 3.4 Functional dissection of the DUF719 and ALPS motif of FAM114A proteins. (**A**) HA-tagged and mitochondrial-relocated DUF719 of FAM114A proteins were tested for their ability to reroute intra-Golgi cargo protein giantin to the mitochondria in HeLa cells. Cells were counterstained with the trans-Golgi marker golgin-245 to visualise the remainder of the Golgi. HeLa cells expressing the GFP-tagged DUF719 (**B**) or the ALPS motif (**C**) of FAM114A proteins with corresponding 'WG' mutants. Cells were stained for the Golgi marker giantin. Scale bars correspond to 10 µm.

In order to explore which other parts of the protein might be contributing to specificity, more truncation mutants were designed. An overexpressed GFP-tagged FAM114A1 truncation (1-410) colocalised partially with giantin but was also localised to various other intracellular punctate structures (see inset, Figure 3.5A) as were observed for the FAM114A2 73-133 truncation mutant. Overexpression of this truncation also caused the Golgi to fragment. Previous work conducted on the orthologous *Drosophila* protein had partially mapped a region containing a predicted coiled-coil as the Rab2 binding site (Gillingham et al., 2014). Moreover, the Rab-binding site of many Rab effectors are predicted coiled-coil regions (Rai et al., 2019). With the aim of possibly removing the Rab2 effector site, the predicted coiled-coil of FAM114A2 was deleted (268-295) and the mutant GFP-tagged and overexpressed. Unlike the full-length protein, the FAM114A2 coiled-coil deletion mutant displayed only limited colocalisation with giantin and instead predominately localised to intracellular puncta adjacent to giantin. Overexpression of the mutant also triggered the vesiculation of the Golgi.

Further work is required to map the Rab2 binding site for the orthologous mammalian FAM114A proteins and to better understand what other factors and what other regions are required for their correct localisation and function. In order to gain a better sense of domain organisation of the FAM114A proteins, the secondary structure and probability of disorder across the protein sequences were predicted using PSIPRED V4.0 and DISOPRED 3 analysis tools respectively. Both FAM114A proteins were predicted to have a disordered N-terminal region of approximately 75-125 amino acids immediately preceding the ALPS/'WG'-motif (Figure 3.5B). The remainder of the proteins were predicted to consist of a series of alpha-helical stretches punctuated by disordered turns. These predictions will serve to facilitate future mapping studies and help understand the biochemical nature of the FAM114A family.

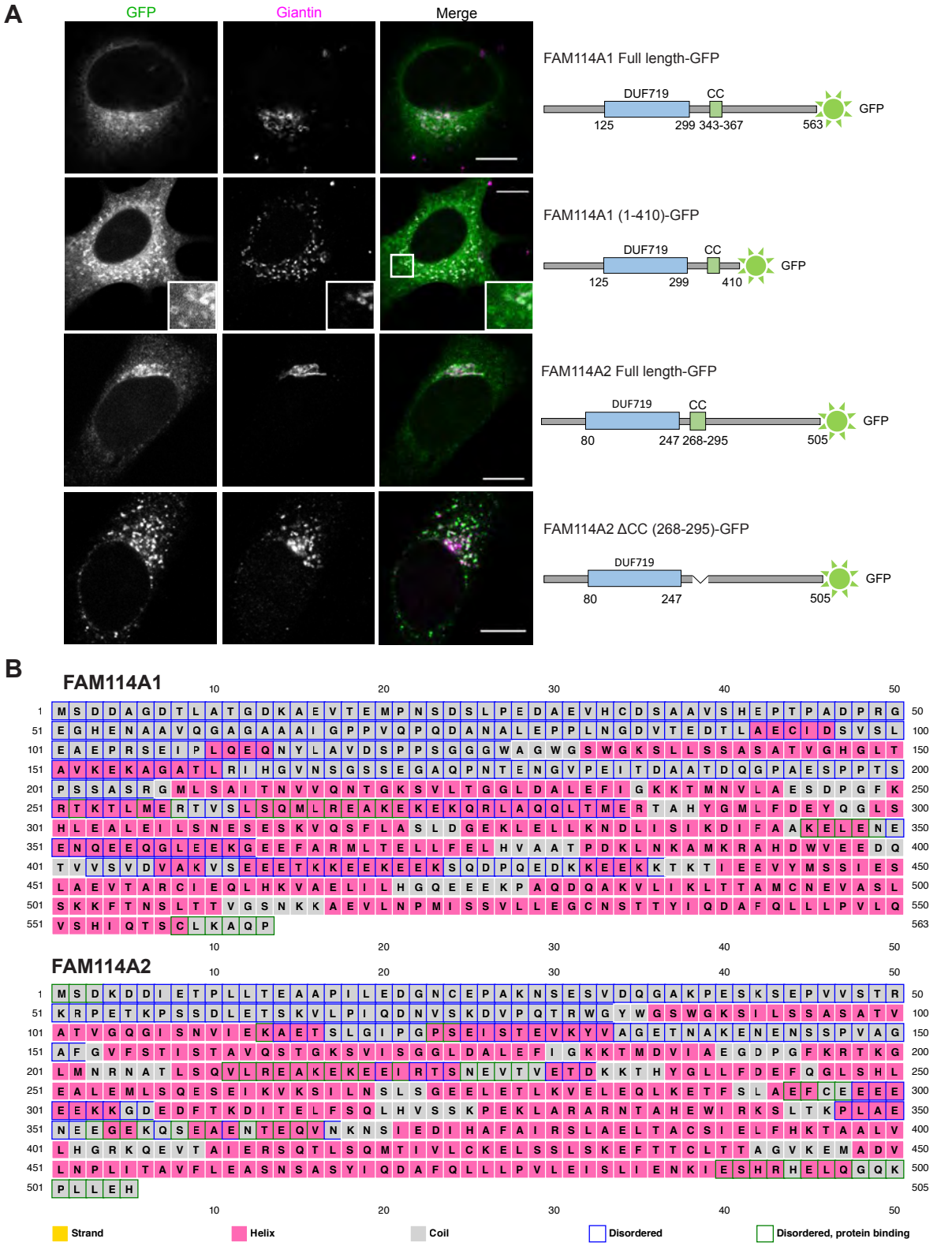


Figure 3.5 Truncation mutants of FAM114A can induce Golgi fragmentation. (**A**) HeLa cells expressing GFP-tagged FAM114A truncation mutants and GFP-tagged wild-type proteins (graphical representation, right). Cells were stained for the Golgi marker giantin. Scale bars correspond to 10 µm. (**B**) A cartoon graphic displaying the predicted secondary structure and disorder of FAM114A proteins generated using the PSIPRED Workbench. The secondary structure was predicted from primary sequences using PSIPRED V4.0. Intrinsically disordered regions and the likelihood of these regions interacting with other proteins was predicted using DISOPRED 3. Key (below), amino acid numbering (above, below and sides).

3.6 Discussion

The application of MitoID to the intra-Golgi golgin tethers GMAP-210 and golgin-84 was successful in identifying FAM114A2, an early-Golgi resident protein of unknown function. Electron microscopy of cells expressing mitochondrial-relocated GMAP-210 and golgin-84 confirms that these tethers are capturing vesicles (diameters of 40-60 nm) as opposed to larger fragments of the Golgi (Wong and Munro, 2014). Therefore, it is possible to infer that FAM114A2 is a vesicle-resident protein since it was detected through proximity-labelling of these tethered vesicles. It is worth noting that while *FAM114A2* displays a near ubiquitous expression profile, expression of ortholog *FAM114A1* appears to be lower and restricted to a more narrow range of tissues and cell types (Thul et al., 2017; Uhlén et al., 2015). Moreover, FAM114A1 is undetectable in HEK293 cells which would explain why it was not detected alongside FAM114A2 in the MitoID screen conducted in this cell type.

MitoID utilises the promiscuous biotin ligase BirA* to biotinylate proximal lysines prior to streptavidin affinity purification and mass spectrometry identification (Roux et al., 2012; Shin et al., 2017). This is particularly useful for the detection of weak or transient interactions which may otherwise be lost using traditional biochemical approaches to detect protein-protein interactions. However, the approach is limited in that labelling requires the substrate lysine residues to be both proximal and available. Therefore, the ligase can only target cytosolic proteins, peripheral membrane proteins (such as FAM114A2) or the cytosolic regions of integral membrane proteins. Amongst the detectable integral membrane proteins, there was a clear a preference for large coiled-coil bitopic proteins such as giantin and golgin-84. The long and extended nature of the cytosolic portions of these proteins increases the likelihood of there being accessible substrate lysines. Furthermore, after purification and trypsin digestion, larger proteins generate more peptides and are consequently more likely to be overrepresented in mass spectrometry analysis. Many novel intra-Golgi vesicle-resident proteins, particularly integral membrane proteins, are likely to be undetectable by MitoID owing to a lack of available cytosolic lysines. This deficiency can be illustrated using the case of GALNT2, a single-pass integral membrane cargo protein which is robustly rerouted by mitochondrial-relocated GMAP-210 and golgin-84 (Wong and Munro, 2014). Despite its robust mitochondrial-relocation by the golgins, it was undetectable by MitoID. GALNT2 has a very short cytoplasmic tail with no lysines (MRRRSR) and is therefore an incompatible substrate for biotinylation. For these reasons, amongst others, the application of MitoID to golgin-84 and GMAP-210 only yielded a handful of relevant candidates.

While it is known that the N-terminus of these golgins is the minimal vesicle-binding region (Wong et al., 2017), the vesicle-derived golgin receptor remains unknown. A combination of biochemical, cell biological and genetic experiments failed to demonstrate that any of the MitoID candidates (Sec23IP, TACC1, TMEM87A, ERGIC-53, giantin) bound directly to the N-terminus of the golgins (data not shown). A future study employing a different proteomic or genetic screen for the receptor will be required to break this impasse.

Mitochondrial-relocated FAM114A proteins were able to reroute early-Golgi but not late-Golgi membrane markers. Confidence can be had that the FAM114A proteins are relocating membranes as all of the proteins successfully tested for relocation were integral membrane proteins (GALNT2, giantin and golgin-84). Mitochondrial-relocation of the FAM114A proteins triggered striking clumping of the mitochondria. Future ultrastructure analysis of these mitochondrial clumps will shed some light on whether the membranes being recruited are vesicles and/or larger Golgi membrane fragments. The relocation of FAM114A proteins also caused membranes of the late-Golgi to fragment and decorate around the mitochondrial clump. This suggests that the FAM114A proteins are not directly interacting with the late-Golgi membranes but rather their repositioning occurs indirectly through intermediate factors and/or membranes. It is conceivable that this ectopic mitochondrial-early-Golgi complex could somehow be acting as a nucleating point for the late-Golgi.

Another striking observation was that mitochondrial-targeted FAM114A2 caused a blockage in ER export. A likely explanation for the inhibition of ER export is that factors required for COPII vesicular transport are sequestered away in the mitochondrial clump. What remains unclear is whether these factors are titrated away due to the capture of anterograde COPII vesicles or retrograde COPI vesicles since many factors cycle between the two carriers. The latter seems more likely since all of the proteins that have been shown to be rerouted by mitochondrial-relocated FAM114A proteins are COPI cargoes. Moreover, ectopic relocation of the cis-Golgi tether GM130, which captures ER-derived COPII-dependent vesicles, does not have the same effect on ER export as FAM114A proteins (Wong and Munro, 2014).

FAM114A proteins have a predicted ALPS motif which was necessary to interact with membranes of the early-Golgi, most likely through absorbing onto highly curved membranes such as that of a vesicle or a cisternal membrane exhibiting high curvature. Future liposome binding assays are required to confirm that the ALPS motif of the FAM114A proteins can interact with highly-curved membranes *in vitro*. The ALPS motif

appears to be highly conserved amongst metazoa; however, in plants the analogous helical region is slightly different. While the helix is still amphipathic, it does not resemble a classical ALPS motif as it has many charged residues on its hydrophilic surface. Interestingly, GMAP-210 also has an ALPS motif and its localised at its N-terminus and has been suggested to be important for its ability to interact with vesicles (Drin et al., 2007; Magdeleine et al., 2016; Roboti et al., 2015). However, what is unclear in both instances is how these proteins interact with this specific subset of highly-curved membranes when many other membranes of high curvature exist throughout the cell. When stripped back to the minimal motif, the ALPS of FAM114A2 localised non-specifically to a range of different membranes *in vivo* (Figure 3.4C), a similar property to what has been reported for other ALPS motifs (Doucet et al., 2015). This suggests that other regions of the FAM114A proteins impart specificity to their membrane-targeting properties. Members of the FAM114A family were initially identified as Rab2 effectors and it is therefore highly plausible that Rab2 at least partially contributes to this specificity (Gillingham et al., 2014; Gillingham et al., 2019). As for other factors which may influence their membrane-targeting, proteomic analysis of FAM114A interactors will be discussed in subsequent chapters.

In summary, MitoID of the intra-Golgi golgins GMAP-210 and golgin-84 identified the intra-Golgi vesicle-resident protein of unknown function FAM114A2. The FAM114A family of proteins specifically interact with membranes associated with the early-Golgi, most likely early intra-Golgi vesicles and/or early-Golgi cisternae exhibiting high-curvature. The highly-conserved predicted ALPS motif of FAM114A proteins is necessary to interact with these highly-curved membranes but other domains are required to impart specificity.

4. Proteomic and biochemical analysis of FAM114A and GOLPH3 family proteins

4.1 Introduction

In the previous chapter, the FAM114A family proteins were identified as novel early intra-Golgi vesicle-resident proteins. FAM114A proteins are predicted to have an ALPS motif which appeared to be necessary for membrane association, most likely through an interaction with highly-curved membranes. However, preliminary findings suggested that regions outside of the ALPS motif were necessary to allow FAM114A proteins to bind specifically to membranes associated with the early-Golgi. It is unclear how these other regions impart specificity as there is very little known about the binding partners of the FAM114A proteins with the exception of Rab2. Furthermore, the FAM114A family are poorly characterised proteins of unknown function and in order to begin to understand their function, more information is required on their binding partners.

4.2 Characterisation of the FAM114A and GOLPH3 family protein interactome

In order to better characterise FAM114A proteins, they were used as baits in GST pulldowns to identify novel protein interactors. GST-tagged FAM114A1 and FAM114A2 (and their corresponding 'WG' mutants) were produced recombinantly in Sf9 cells using a baculoviral expression system. The GST-tagged FAM114A proteins were purified and then used to isolate binding partners from HEK293T cell lysate which were identified by mass spectrometry and subject to quantitative analysis. When compared to the GST control, only a handful of proteins were enriched in the FAM114A1-GST sample including both Rab2 isoforms, the uncharacterised protein C16orf13, a component of the ATP synthase complex (ATP5C1), translational proteins (CARS, EFF1B2 and EEF1G) and glycan-modifying enzyme HS2ST1 (Figure 4.1A, Supplementary Table 8.1). When compared to the GST control, FAM114A2 also specifically interacted with C16orf13 and components of the ATP synthase complex (ATP5C, ATP5D) and translational proteins (CARS, EEF1A1, EFF1B2 and EEF1G; Figure 4.1B, Supplementary Table 8.2). A noticeable difference was that FAM114A2 also pulled down a number of COPI cargo proteins including a variety of glycan-modifying enzymes and p24 cargo receptors (Figure 4.1B, Supplementary Table 8.2). Furthermore, FAM114A2 also pulled down some components of the oligosaccharyltransferase (OST) complex (DDOST, RPN2), lipid-modifying enzymes (ALDH3A2, TMEM55A, TMEM55B), lipid transporters (SLC27A1, SLC27A4), the G protein β1-subunit (GNB1) and microsomal GST-3 (MGST3). There was very little difference in protein interactors when comparisons were made between wild-type FAM114A proteins

and their corresponding 'WG' mutant (Figure 4.1, right panels; Supplementary Tables 8.3, 8.4). This provides support for the idea that mutation of the tryptophan residues of the 'WG' repeat ablate a protein-lipid interaction (the ALPS motif absorbing onto highly-curved membranes) rather than a specific protein-protein interaction. Furthermore, it suggests that mutagenesis does not trigger gross misfolding of either protein as they retain their ability to interact with their binding partners and there is not an increased binding of chaperones.

FAM114A2 was able to pulldown a number of COPI cargo proteins. Since the FAM114A family are peripheral membrane proteins and many of the cargo proteins are single pass transmembrane proteins, it is highly likely they are interacting with the cytoplasmic tails of the cargo. This is similar to the COPI adaptors GOLPH3 and GOLPH3L which have been proposed to simultaneously interact with the cytoplasmic tails of cargo and the COPI coat (Eckert et al., 2014; Tu et al., 2008; Tu et al., 2012). While the initial work carried out on the yeast ortholog *Vps74* identified (F/L)(L/I/V)XX(R/K) as the recognition motif in cytoplasmic tails (Tu et al., 2008), the recognition motif for GOLPH3 and GOLPH3L is poorly defined. Moreover, until recently there had only been a handful of reported clients (Chapter 1, Table 1.4.1). In order to make proteomic comparisons between the FAM114A and GOLPH3 family proteins and to expand the GOLPH3 family client list, GOLPH3 and GOLPH3L were used as baits in GST pulldowns to identify novel interactors. GST-tagged GOLPH3 family proteins were produced recombinantly in bacteria, purified and used to isolate binding partners from HEK293T cell lysate. When the samples for GOLPH3 and GOLPH3L were compared to the GST control, it revealed a significant enrichment for many of the subunits of the COPI coat, over a hundred COPI cargo proteins and many other proteins (Figure 4.2A, Table 4.2.1, Supplementary Table 8.5). Shown in Table 4.2.1 are the top 50 GOLPH3 and GOLPH3L interactors.

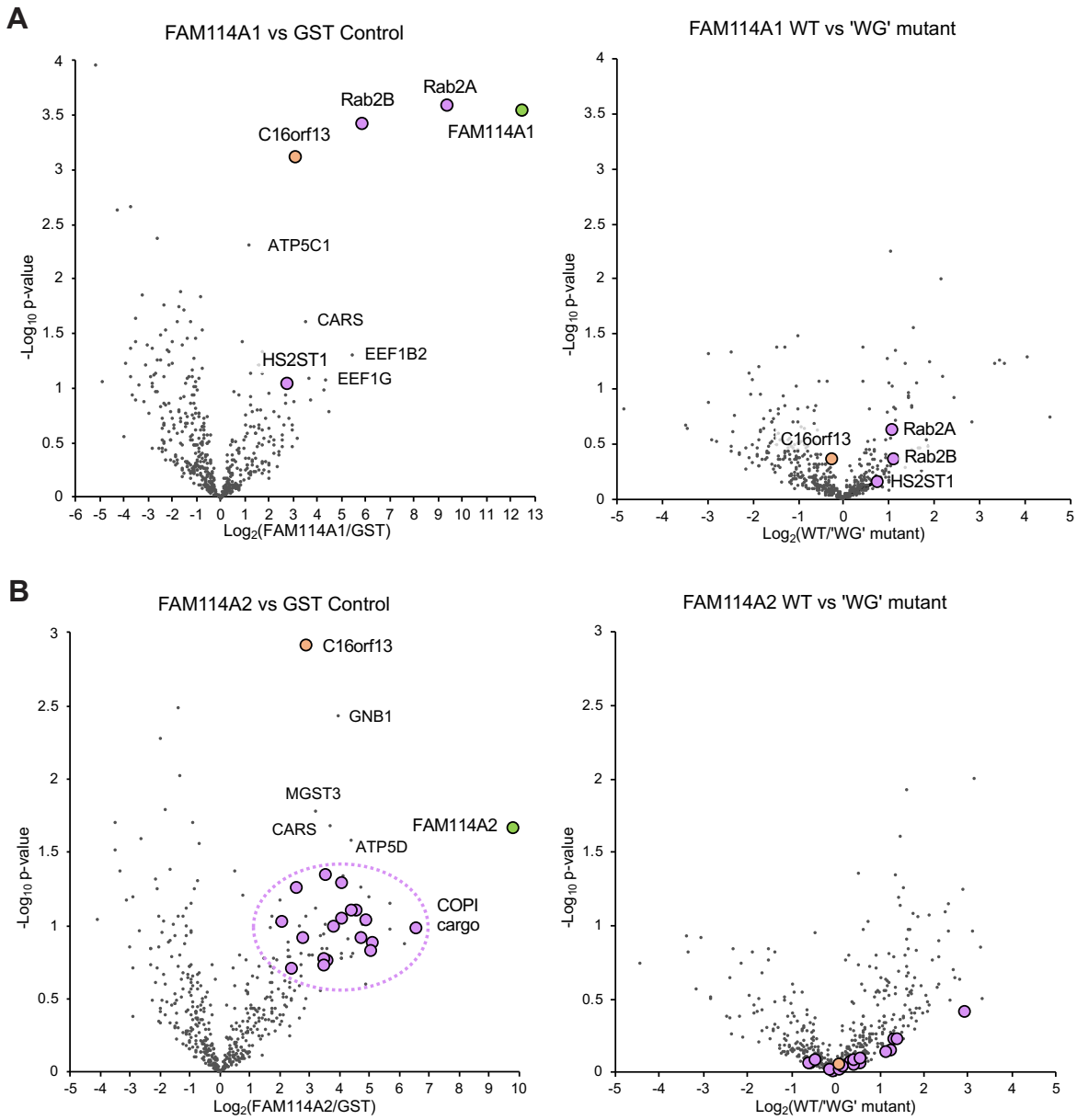


Figure 4.1 Proteomic analysis of FAM114A proteins by GST affinity chromatography and mass spectrometry. Volcano plots comparing spectral intensity values generated from GST pulldowns from HEK293T cell lysate with GST-tagged FAM114A1 (**A**) or FAM114A2 (**B**) vs a GST control or the corresponding 'WG' mutant. -log P values were generated from t-tests. Bait proteins are highlighted in green, known vesicular proteins in magenta and C16orf13 in orange. The highlighted proteins are consistent across adjacent plots. Data was analysed using the Perseus platform and represent a triplicate of repeats.

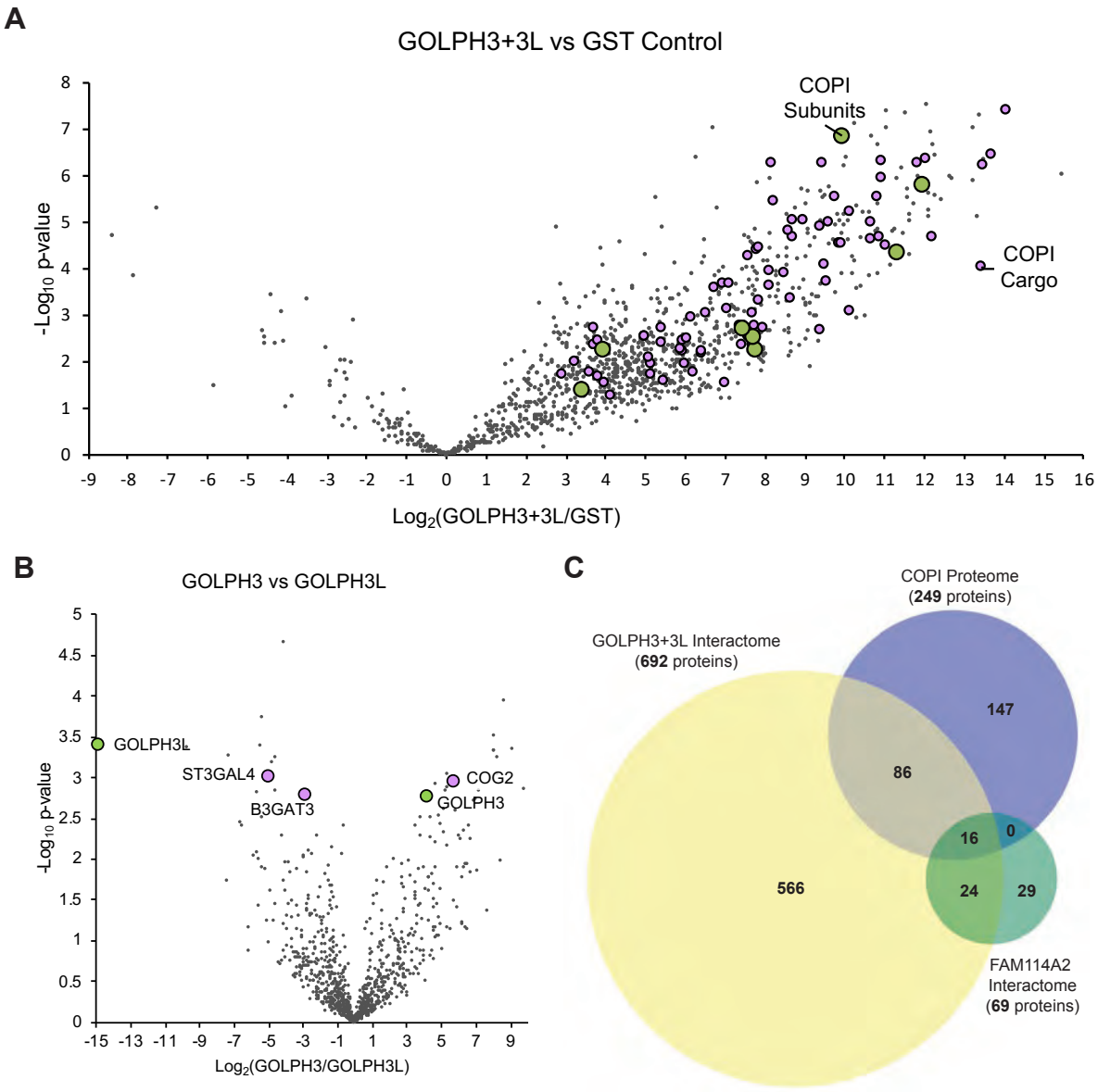


Figure 4.2 GOLPH3 and GOLPH3L simultaneously pulldown the COPI coat and a host of COPI vesicular cargo proteins. Volcano plots comparing spectral intensity values generated from GST pulldowns from HEK293T cell lysate with GST-tagged GOLPH3 and GOLPH3L combined vs a GST control (**A**) or GOLPH3 vs GOLPH3L (**B**). -log P values were generated from t-tests. Highlighted in green are subunits of the COPI coat (**A**) or bait proteins (**B**) and known COPI cargo or COPI-associated proteins (non-exhaustive) in magenta. Data was analysed using the Perseus platform and represent a triplicate of repeats. (**C**) A Venn diagram comparing the top hits from the FAM114A2 GST pulldown data set (69 proteins with lowest p-value - FAM114A2 vs GST) the combined GOLPH3 and GOLPH3L GST pulldown data set (proteins for which $P \leq 0.05$ GOLPH3+3L vs GST) and a reference COPI proteome data set (Figure 5A-C, proteins with a mean $\log_2$ SILAC ratio of > 0, sample vs control; Adolf et al., 2019).

Table 4.2.1

Gene names	-Log_{10} p-value	Difference Log_2(GOLPH3+3L/GST)
DARS2	6.02	15.44
C3orf58	7.38	14.07
ZFPL1	6.46	13.70
COA3	6.35	13.49
GALNT2	6.20	13.48
MOGS	4.05	13.46
LRRC59	7.31	13.38
DNAJC19	5.12	13.32
PAM16	5.90	13.22
SUN2	7.04	13.22
ABHD12	5.93	12.67
TMEM206	6.00	12.61
SPTLC2	5.51	12.43
NOC4L	6.44	12.28
PIGS	5.57	12.27
PEX16	6.68	12.23
UGT8	4.67	12.20
PIGK	6.93	12.15
STX12	5.64	12.13
RIF1	5.11	12.12
TMEM201	5.27	12.12
MAN1B1	6.36	12.04
NDUFB10	7.52	12.03
COPB1	5.80	11.95
FAM3C	5.77	11.92
TEX10	6.31	11.86
TMEM192	6.00	11.85
QPCTL	6.24	11.84
FAF2	5.38	11.74
LETMD1	5.03	11.64
ESYT2	4.28	11.63
BRI3BP	5.16	11.60
SPTLC1	5.70	11.59
TAMM41	5.55	11.56
OPA1	7.36	11.52
AFG3L2	4.77	11.45
NOP14	6.16	11.44
PHB	5.96	11.29
COPG2	4.36	11.29
GDAP1	4.83	11.28
TOMM70A	4.80	11.21
NOC2L	4.07	11.15
SIGMAR1	7.41	11.07
LETM1	4.50	11.06
GALNT7	4.49	11.03
OGFOD3	7.02	11.03
QSOX2	5.95	10.92
B4GAT1	6.31	10.91
GOSR1	4.67	10.89

-Log_{10} p-values were generated from a Student's t-test comparing protein spectral intensities. The table corresponds to a truncated form of the data presented in Figure 4.2A and Supplementary Figure 8.5. COPI cargo (magenta) and coat subunits (green).

GOLPH3 and GOLPH3L share 80% sequence similarity so they were compared to one another to see if there were any differences in their binding specificities. There were very few differences between the GOLPH3 family members; of note, GOLPH3 specifically pulled down a subunit of the conserved oligomeric Golgi complex (COG2) whereas GOLPH3L specifically pulled down the glycan-modifying enzymes ST3GAL4 and B3GAT3 (Figure 4.2B, Supplementary Table 8.6). Recently, Adolf and colleagues profiled the proteome of COPI vesicles purified from semi-permeabilised HeLa cells (Adolf et al., 2019). Since both FAM114A2 and GOLPH3 family proteins pulled down many putative COPI cargo proteins, this dataset served as a useful reference COPI proteome for comparisons. A three-way comparison between FAM114A2 interactors, GOLPH3+3L interactors and the reference COPI proteome was made (Figure 4.2C, Supplementary Table 8.7). Of the 249 proteins in the COPI proteome, 102 (41%) featured in the GOLPH3+3L interactome. There were no proteins which were present in the COPI proteome which exclusively bound to FAM114A2 over GOLPH3+3L. It is worth noting that the COPI proteome is limited as many proteins which are expected to be COPI cargo proteins (e.g. known Golgi-resident glycan-modifying enzymes) are found in the FAM114A2 and GOLPH3+3L interactomes but not in the COPI proteome. Regardless, the data suggests that GOLPH3+3L are highly promiscuous COPI adaptors while FAM114A2 interacts with a more select pool of cargo. As a consequence, while there are plenty of cargo proteins exclusive to GOLPH3+3L, there were no cargo proteins exclusive to FAM114A2. In summary, GST affinity chromatography proteomic analysis of GOLPH3+3L family proteins has potentially revealed over a hundred of novel COPI cargo proteins, many of which are glycan-modifying enzymes. Proteomic analysis of FAM114A proteins has confirmed an interaction between FAM114A1 and Rab2 and has also revealed that FAM114A2 can bind *in vitro* to a selection of COPI cargo proteins. The analysis also confirmed an *in vitro* interaction of the GOLPH3 family, but not the FAM114A family, with the COPI coat. The analysis has also shown that FAM114A1 and FAM114A2 can interact with another novel protein, C16orf13, a protein that requires attention in future studies.

4.3 FAM114A2 binds directly to the cytoplasmic tails of glycan-modifying enzymes

As alluded to previously, as FAM114A2 is a peripheral membrane protein and since the interacting cargo proteins are single-pass transmembrane proteins, FAM114A2 is likely to be interacting with these proteins through their cytoplasmic tails. Most of the cargo proteins are both Golgi-resident glycan-modifying enzymes and type II transmembrane proteins which exist as homo- or heterodimers. Therefore, it is not clear if the proteins are interacting with the cytoplasmic tails directly or indirectly as a result of heterodimerisation with another enzyme or via another intermediary. Recently, the cytoplasmic tail and TMD of the type II transmembrane protein sucrase-isomaltase was fused to GFP for use as a plasma membrane reporter (Liu et al., 2018). In order to validate the interaction between FAM114A2 and these Golgi enzymes *in vitro*, a series of chimeric reporters were generated in which the cytoplasmic tails of the enzymes replaced that of sucrase-isomaltase with its catalytic domain removed (Figure 4.3A). The chimeras were fused at their C-terminus to a GFP-FLAG tag and overexpressed in HEK293T cells. A chimera bearing the cytoplasmic tail of one of the strongest FAM114A2 interactors, GALNT2, was pre-purified from HEK293T cell lysate by FLAG affinity chromatography. GST-tagged FAM114A proteins were purified in Sf9 cells and used as baits to test direct binding to the FLAG-purified chimera in a GST pulldown in the presence of detergent. Only FAM114A2, but not FAM114A1 or the GST negative control, bound directly to the GALNT2 cytoplasmic tail chimera (Figure 4.3B). Next, several chimeras were generated using cytoplasmic tails from glycan-modifying enzymes from a broad spectrum of enzyme families. These chimeras were overexpressed in HEK293T cells and the lysate was used to test binding to GST-tagged FAM114A2 produced in bacteria. FAM114A2 showed a range of binding affinities for the different tails, binding strongly to the tail chimera of GALNT2, GALNT12, B3GALT6, MGAT2, GOLIM4 and MANEAL (Figure 4.3C). In contrast, FAM114A2 bound weakly to CHSY1, B3GAT3, GOLM1 and ST6GAL1 and not at all to MGAT5, CASC4, TPST2 or the sucrase-isomaltase control.

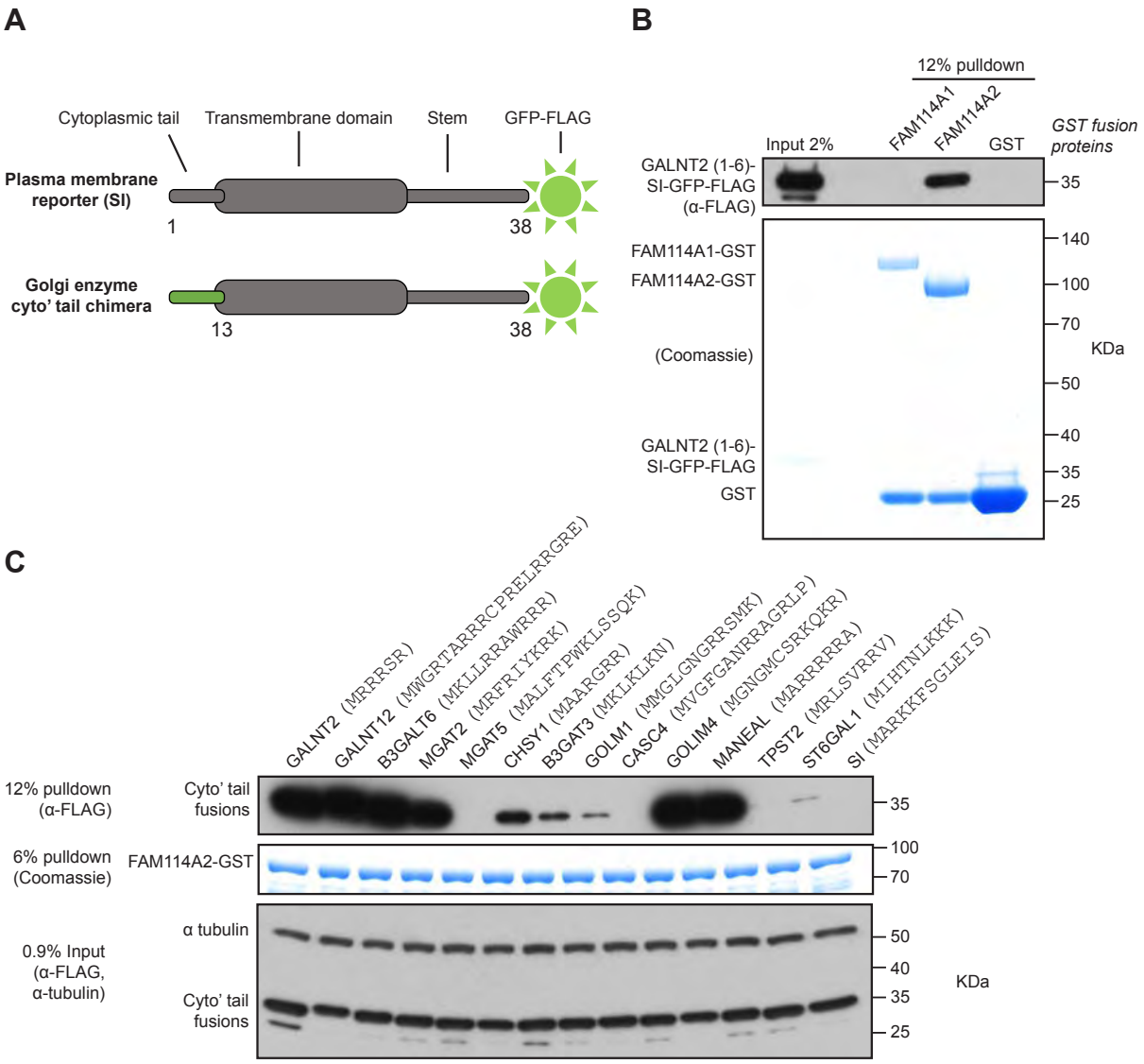


Figure 4.3 FAM114A2 interacts directly with the cytoplasmic tails of Golgi enzymes. (**A**) A cartoon schematic showing the Golgi enzyme cytoplasmic tail chimeras used in binding experiments. Chimeras are type II transmembrane proteins. SI, sucrase-isomaltase. (**B**) A Coomassie stain and immunoblot from a GST pulldown testing direct binding between a GALNT2 cytoplasmic tail chimera (overexpressed in HEK293T cells and pre-purified from lysate by FLAG affinity chromatography) and GST-tagged FAM114A proteins purified from Sf9 cells. N=1. (**C**) A Coomassie stain and immunoblots from a GST pulldown testing interactions between a host of different cytoplasmic tail chimeras and GST-tagged FAM114A2 expressed in bacteria. Gene names associated with the cytoplasmic tails are given immediately adjacent to blots and the corresponding tail sequences according to Uniprot are given in brackets. N=2.

A theme amongst the cytoplasmic tails of the strongest binders to FAM114A2 was the presence of membrane-proximal poly-arginine stretches. In order to test if this was of significance for FAM114A2 binding, a couple of new enzyme tail chimeras were generated which shared this feature (UXS1 and GCNT1). Moreover, the poly-arginine stretch of GALNT2 was subject to lysine-scanning mutagenesis while arginines of B3GALT6 and MANEAL were also targeted by mutagenesis. Conversely, a membrane-distal poly-arginine stretch was added to the tail of sucrase-isomaltase through missense mutagenesis of lysine residues. Across the mutant chimeras, mutagenesis of arginines to lysine only had very mild negative effects on the binding affinity to FAM114A2 whereas introduction of a membrane-distal poly-arginine stretch to sucrase-isomaltase failed to confer any binding (Figure 4.4A). A re-evaluation of the tails of the strongest binders suggested that there was possibly a requirement for several membrane-proximal positively-charged residues for FAM114A2 binding. More precisely, many of the tails had at least three positively-charged residues in the six amino acid window adjacent to the membrane. It is worth noting that all binding experiments are conducted with TMD-containing chimeras in the presence of detergent. It is therefore likely that the detergent micelle is somewhat mimicking a lipid bilayer *in vitro*. To test the significance of membrane/micelle-proximal positive charges for FAM114A2 binding *in vitro*, the tail of sucrase-isomaltase was subject to a membrane-proximal insertion or intercalation of three arginines, lysines or histidines (not expected to be protonated in the cytosol). The membrane-proximal addition of three positively-charged residues, but not three histidines, was sufficient to confer FAM114A2 binding to the chimeras (Figure 4.4B). Similarly, mutagenesis of a membrane-proximal arginine to alanine or histidine in the tail of GCNT1 caused a marked reduction in FAM114A2 binding. In an attempt to increase the distance between the membrane and the basic residues of the tail of GALNT2, alanine residues were inserted proximal to the membrane. The GALNT2 mutant chimeras appeared to exhibit a decrease in FAM114A2 binding in a manner dependent on the length of alanine extension. However, the stability of the chimera decreased slightly with alanine extension so it is not clear if this represented a true reduction in binding affinity. Moreover, it must be noted that alanine is a hydrophobic residue so it may reside in the membrane/micelle and negate the intended extension of the cytoplasmic tail.

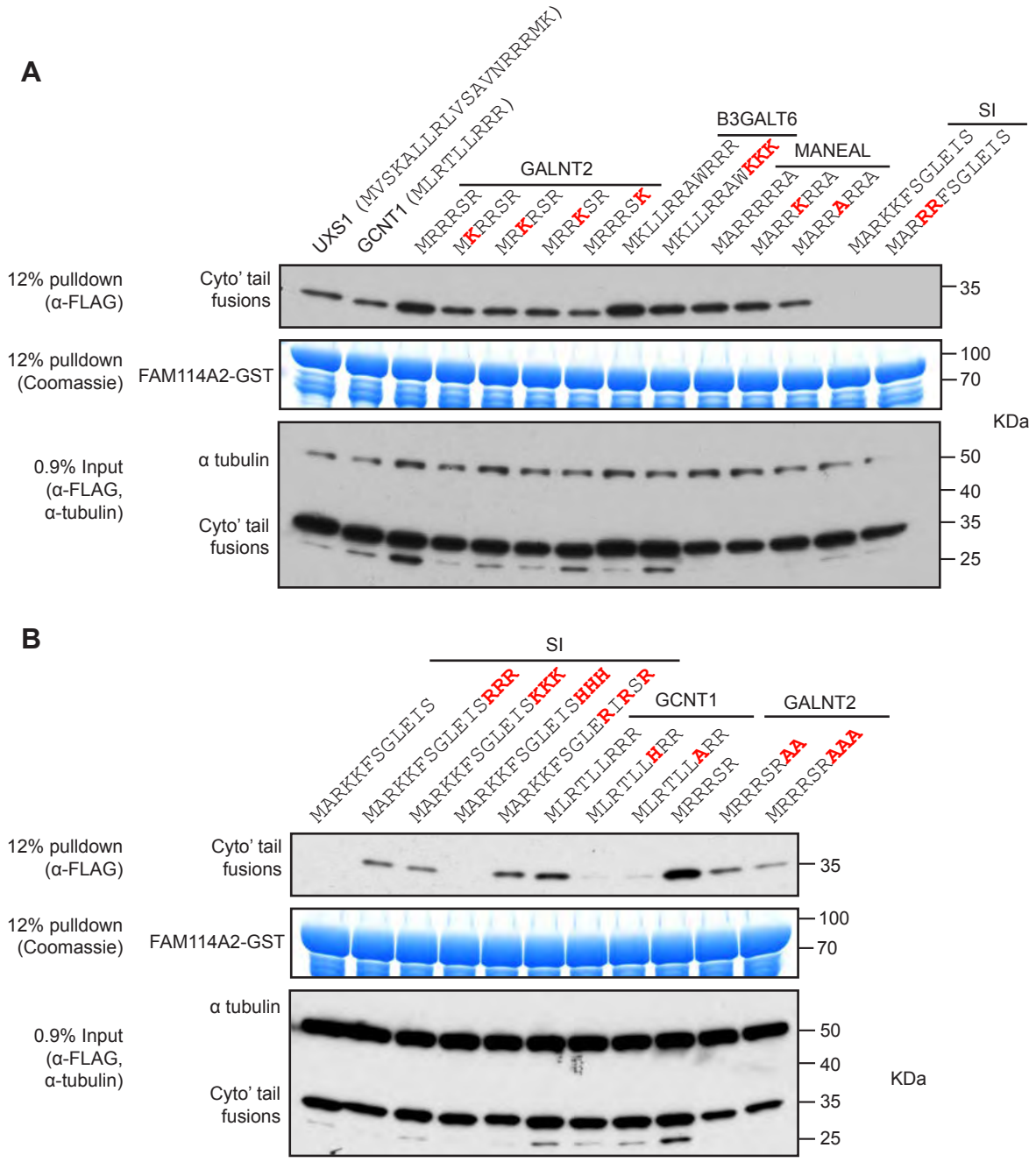


Figure 4.4 FAM114A2 recognises membrane-proximal basic residues in the cytoplasmic tails of type II transmembrane proteins. (**A**+**B**) Coomassie stains and immunoblots from a GST pulldown testing the ability of bacterial expressed GST-tagged FAM114A2 to pulldown different cytoplasmic tail chimeras from HEK293T cell lysate. The cytoplasmic tail sequences are given in brackets or immediately adjacent to blots. Cytoplasmic tail sequences are given according to Uniprot. Red letters indicate missense or insertion mutant residues resulting from targeted mutagenesis. Gene names are given immediately adjacent to blots (UXS1 and GCNT1) or above sequences. N=1.

4.4 Mapping the cytoplasmic tail binding site of FAM114A2

To map to the site on FAM114A2 that interacts with the cytoplasmic tails of the Golgi enzymes, a series of GST-tagged truncation mutants were designed. The truncation mutants were purified from Sf9 cells and used as baits to test direct binding to the FLAG-purified GALNT2 cytoplasmic tail chimera in a GST pulldown. In an initial attempt to map the binding site, only FAM114A1, the N-terminus of FAM114A2 (1-79) and the GST negative control failed to bind to the GALNT2 tail chimera (Figure 4.5A). The 80-347 FAM114A2 truncation mutant pulled down relatively less of the chimera corresponding to poorer expression of the bait as opposed to a lower binding affinity. On a subsequent attempt to map the binding site, a second collection of truncation mutants were designed with a focus on mapping a binding site in the central portion and at the C-terminus of FAM114A2. Two binding sites were identified at positions 134-347 and at the extreme C-terminus between residues 442-505 (Figure 4.5B). The central binding site is over 200 amino acids and encompasses part of the DUF719 and the coiled-coil region (Figure 4.5C). Further studies are required to map this binding site more closely.

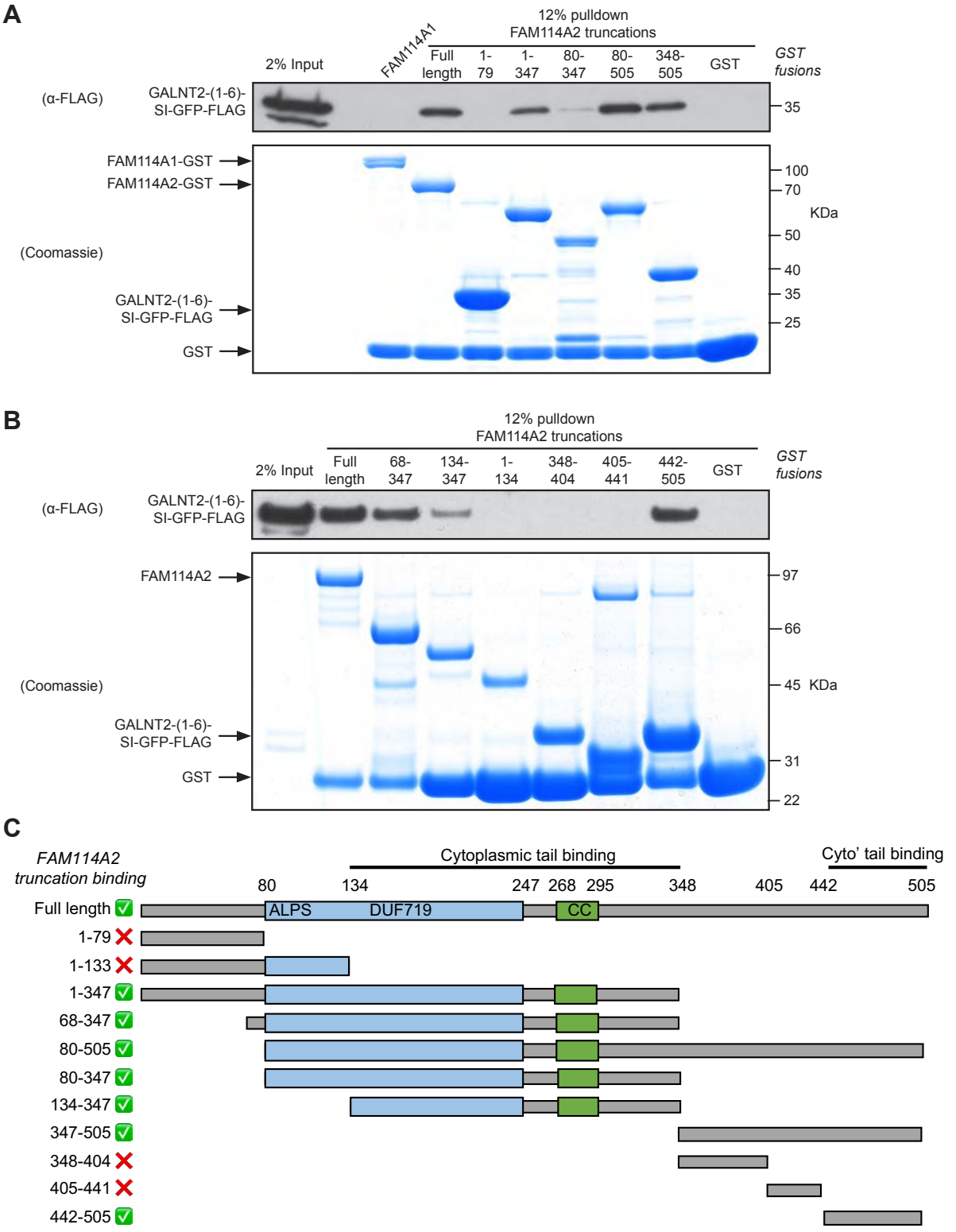


Figure 4.5 Mapping the interaction between FAM114A2 and the cytoplasmic tails. (**A**+**B**) Coomassie stains and immunoblots from a GST pulldown testing direct binding between a GALNT2 cytoplasmic tail chimera (overexpressed in HEK293T cells and pre-purified from lysate by FLAG affinity) and GST-tagged FAM114A truncation mutants purified from Sf9 cells. N=1. (**C**) A schematic displaying the different truncations and their ability to bind to the cytoplasmic tail chimera. CC denotes a predicted region of coiled-coil, amphipathic lipid packing sensor (ALPS).

4.5 GOLPH3 recognises cytoplasmic tails with a total charge of ≥4+

GST affinity chromatography using GOLPH3 and GOLPH3L as baits potentially identified over a hundred novel clients. In order to validate some of these novel clients and to elucidate the GOLPH3 consensus motif, the cytoplasmic tail chimeras were tested for their ability to bind to GST-tagged GOLPH3 purified from bacteria. GOLPH3 was able to pulldown all of the tail chimeras initially tested with the exception of MGAT5 and sucrase-isomaltase (Figure 4.6). A relatively small amount of the GALNT3 cytoplasmic tail chimera was pulled down by GOLPH3 but that was reflective of the instability of the chimera as opposed to low binding affinity. No clear consensus motif was immediately visible from the selection of tail chimeras which bound to GOLPH3; however, binding appeared to be charge dependent. When accounting for the positive charge of the amino terminus and that histidine is not protonated at a cytosolic pH of 7.4, cytoplasmic tails with a predicted charge of 4+ or higher bound to GOLPH3 whereas those with a predicted charge of 3+ or less failed to bind convincingly. It is worth noting that a serine residue in the cytoplasmic tail of sucrase-isomaltase has been reported to be phosphorylated which would lower the predicted charge (Keller et al., 1995).

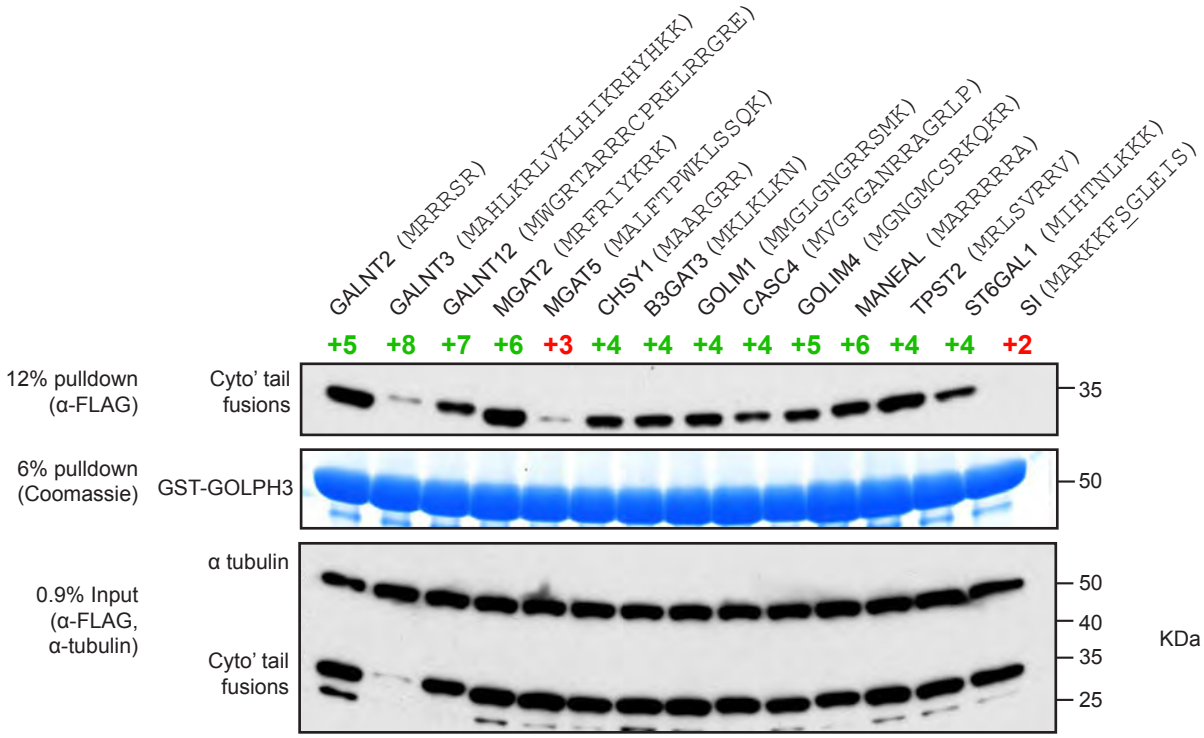


Figure 4.6 GOLPH3 interacts with the cytoplasmic tails of Golgi enzymes. A Coomassie stain and immunoblots from a GST pulldown testing interactions between a host of different cytoplasmic tail chimeras and GST-tagged GOLPH3 expressed in bacteria. The predicted charges of the tails are given immediately adjacent to blots. Predicted charges less than or equal to +3 are given in red, otherwise in green. Total charge values are based on predictions at a cytosolic pH of 7.4 and include the positive charge of the amino terminus. Note that the tail of SI is predicted to be phosphorylated at the serine at position 7. Gene names associated with the cytoplasmic tails are given above and the corresponding tail sequences according to Uniprot are given in brackets. N=2.

To interrogate this charge dependency rule, a handful of chimeras were generated using enzyme tails which do not meet the predicted charge threshold. None of the enzyme tail chimeras with charges of less than 4+ bound to GOLPH3 (FUT3, GALNT4, MGAT1 and TPST1) while those that satisfied the threshold did bind (GALNT2 and EXTL3, Figure 4.7A). An alternative strategy was taken in which site-directed mutagenesis of the negatively-charged residues of the tail of sucrase-isomaltase was conducted to see if it was sufficient to confer GOLPH3 binding. As expected, a tail bearing a phosphomimetic mutation of serine 7 of sucrase-isomaltase had no effect on its ability to bind GOLPH3. Single alanine point mutations in either negatively charged residue alone was insufficient to meet the predicted charge threshold or confer GOLPH3 binding. However, when both negatively charged residues were simultaneously targeted, this was sufficient to meet the charge threshold of 4+ and was sufficient to confer GOLPH3 binding. One concern is that mutagenesis of the cytoplasmic tail of sucrase-isomaltase may be inadvertently destroying a charge-independent GOLPH3 evasion motif. To overcome this, the tail of sucrase-isomaltase was subject to membrane-proximal insertion of basic residues so that the predicted charge threshold was satisfied without the removal of native residues. Only insertion mutants with charges of 4+ or over (+RRR, +ARR, +KKK) and not those with charges of less than 4+ (+AAR, +AAA, +HHH) exhibited binding to GOLPH3 (Figure 4.7B). The FAM114A2 cytoplasmic tail recognition criteria has a positional dependency in relation to the membrane/micelle. In order to explore the possibility of positional dependency in the recognition by GOLPH3, the tail of GALNT2 was subject to membrane proximal alanine insertion mutagenesis to extend the cytoplasmic tail. The cytoplasmic tail extension of the chimeras had subtle negative effects on GOLPH3 binding (Figure 4.7B). In summary, GOLPH3 recognised the short cytoplasmic tails of type II transmembrane proteins with a predicted total charge of 4+ or higher. This imparts promiscuity into the cargo selection of GOLPH3 which may allow it to operate as major intra-Golgi COPI adaptor.

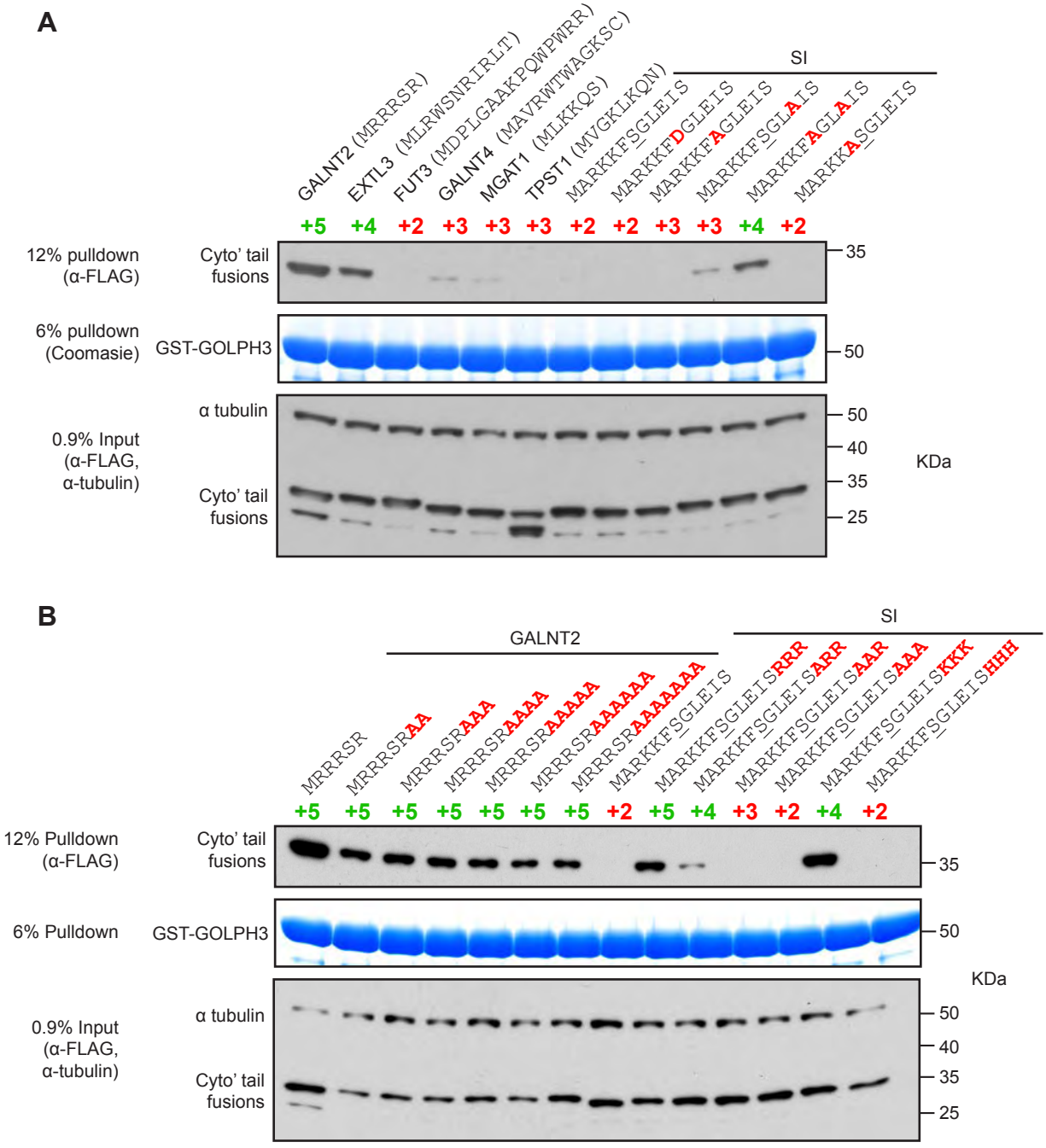


Figure 4.7 GOLPH3 preferentially binds to cytoplasmic tails of type II transmembrane proteins with a charge of +4 or above. (**A**+**B**) Coomassie stains and immunoblots from GST pulldowns testing the ability of bacterial expressed GST-tagged GOLPH3 to pulldown different cytoplasmic tail chimeras from HEK293T cell lysate. The predicted charges of the tails are given immediately adjacent to blots. Predicted charges less than or equal to +3 are given in red, otherwise in green. Total charge values are based on predictions at a cytosolic pH of 7.4 and include the positive charge of the amino terminus. Note that the tail of SI is predicted to be phosphorylated at the serine at position 7 (underlined). The cytoplasmic tail sequences are given in brackets or above charges. Cytoplasmic tail sequences are given according to Uniprot. Red letters indicate missense or insertion mutant residues resulting from targeted mutagenesis. Gene names are given above charges or above sequences. (**A**) N=2. (**B**) N=1.

4.6 Discussion

FAM114A2 was shown to pulldown a host of COPI cargo proteins, including a number of glycan-modifying enzymes. The enzymes belonged to a broad range of different families including those associated with N-linked glycosylation (OST subunits, MOGS, MANEAL, MGAT2) and those associated with O-linked glycan modifications such as O-mannosylation (POMGNT1), proteoglycan synthesis (UXS1, FAM20B, CHPF2) and mucin-type glycosylation (GALNT2, GALNT3, GALNT12). A common theme was that many of the enzymes operated in the early steps of their glycan-modification pathways and have been shown to, or are predicted to reside in, the ER or early-Golgi compartments. Taking the N-linked enzymes as an example; OST and MOGS reside in the ER, MANEAL in the cis-Golgi and MGAT2 in the medial-Golgi (Helenius and Aebi, 2004; Kornfeld and Kornfeld, 1985). As an example of the early-acting nature of some these enzymes, the GALNTs and the OST complex catalyse the initial conjugation of sugar moieties to the polypeptide in O-linked mucin-type and N-linked glycosylation respectively. Furthermore, FAM114A2 pulled down a number of p24 cargo receptors (TMED2, TMED10) which cycle between the ER and the Golgi (Gomez-Navarro and Miller, 2016). Together, this supports the idea from the previous chapter that FAM114A proteins specifically reside on intra-Golgi vesicles recycling from early-Golgi cisternae and possibly even Golgi to ER vesicles.

The FAM114A proteins specifically pulled-down various translational proteins, mitochondrial proteins and lipid-modifying enzymes. Many of these proteins do not share the Golgi-localisation profile of the FAM114A proteins and therefore these interactions are less likely to represent a true interaction i*n vivo*. For this reason, the investigation of these potential interactions was not prioritised in this thesis. Both FAM114A proteins pulled-down the protein of unknown function C16orf13 very convincingly. Interestingly, an antibody raised against C16orf13, also known as methyltransferase-like 26, localises to the Golgi (Atlas Antibodies). Future studies are required to validate the interaction of C16orf13 with the FAM114A proteins and to understand the functional significance of this interaction.

While FAM114A1 and FAM114A2 shared some similarities in how they behaved in a cellular context (Chapter 3), *in vitro* they behaved very differently. While FAM114A2 bound to numerous COPI cargo proteins, FAM114A1 only bound to HS2ST1. In contrast, while FAM114A1 pulled-down both Rab2 isoforms, FAM114A2 pulled down neither. The basis for this discrepancy is not immediately clear and requires further detailed structure-function analysis. A notable discrepancy between the FAM114A family proteomic data sets compared to that of the GOLPH3 family was that the COPI coat was absent. Since FAM114A2 is a vesicle-resident protein which has been shown to bind directly to the tails of cargo, an attractive hypothesis might be that FAM114A2 functions as a novel COPI adaptor

for early intra-Golgi vesicles. This model is less convincing without evidence of an interaction with the COPI coat *in vitro*. However, it is plausible that coatomer binding only occurs once FAM114A2 is docked onto membranes and therefore this interaction may be missed with the binding assay in its current format. A variety of failed attempts were made to demonstrate an interaction between the FAM114A proteins and subunits of the COPI coat *in vitro* which have not been reported in this thesis. The function of FAM114A proteins as COPI adaptors will be explored in detail in the following chapters.

When compared statistically to the GST negative control, GOLPH3 and GOLPH3L were able to specifically pulldown nearly 700 proteins. Many of these proteins were COPI coat subunits or bona fide COPI cargo and therefore appeared to be functionally relevant based on current models of the function of GOLPH3 family proteins (Rizzo et al., 2019; Schmitz et al., 2008; Tu et al., 2008). However, validation is required to ascertain how many of these cargo proteins bind directly to GOLPH3. Moreover, many proteins were integral membrane proteins of non-Golgi compartments including peroxisomes, mitochondria, ER and the nuclear membrane amongst others (Table 4.2.1, Supplementary Table 8.5). It seems highly unlikely that the GOLPH3 proteins are interacting with all of these proteins *in vivo* but rather it represents an artefact of the *in vitro* modality of the experiment. Nevertheless, these spurious interactions may still be revealing of the properties and function of GOLPH3 *in vivo*. In cells, GOLPH3 proteins are specifically recruited to the trans-Golgi via PtdIns4P and once docked onto the membrane, they can sample the cytoplasmic tails of their Golgi-resident clients (Schmitz et al., 2008; Tu et al., 2008; Wood et al., 2009). Interrogation of the GOLPH3 consensus motif suggested that there was simple charge dependency in the cytoplasmic tails of the clients. *In vivo* this may allow GOLPH3 family proteins to bind promiscuously to a vast range of Golgi-resident clients and serve as a master COPI adaptor for intra-Golgi vesicles. In effect, GOLPH3 proteins may act as general gatekeepers at the trans-Golgi to recycle back a broad array of Golgi (and possibly even ER) residents which may have ventured too far along the early-secretory pathway. *In vivo*, the PtdIns4P coincidence detection ensures that this role as a promiscuous filter is confined to the trans-Golgi. However, *in vitro*, in the presence of detergent this coincidence detection is lost and with it so is the specificity by which GOLPH3 proteins interacts. This allows the GOLPH3 proteins to interact spuriously with non-Golgi integral membrane proteins through recognition of positive charges adjacent to their TMDs.

Previous reports have shown that GOLPH3 can bind directly to soluble peptides corresponding to the cytoplasmic tails of clients *in vitro* (Ali et al., 2012; Rizzo et al., 2019). The GOLPH3 binding experiments shown in this thesis were conducted using the tails in the context of TMD-containing chimeras and in the presence of detergent. However, initial attempts to recapitulate binding with the tails in the absence of a TMD and detergent have failed (data not shown). An attractive mechanistic model for GOLPH3 is that it exists in an autoinhibited conformation in the cytosol which is relieved once docked onto trans-Golgi

membranes. This autoinhibition would serve to stop the GOLPH3 family members from non-specifically sampling the positive charges of proteins in the cytosol or on membranes other than the trans-Golgi. This would ensure that GOLPH3 does not act too early in the Golgi stack. *In vitro*, detergent micelles may be sufficient to mimic a trans-Golgi membrane and alleviate the autoinhibition of GOLPH3 proteins, allowing them to bind to the tails. Further binding experiments are required to explore this potential mechanism of autoinhibition and to recapitulate direct binding of GOLPH3 with the cytoplasmic tail peptides *in vitro*.

In a recent report attempting to characterise the GOLPH3 recognition motif, the authors tentatively suggest that GOLPH3 recognises tails with an LxxR/K motif between 3-6 amino acids downstream of at least one positively charged residue (H/R/K [X]2-5 LxxR/K), a motif they term the dual GOLPH3 binding (DGB) motif (Rizzo et al., 2019). The authors do not disregard the potential for other recognition motifs and the DGB is not entirely incompatible with the ≥4+ charge dependency rule suggested in this thesis. Nearly all of the tails tested by the authors gave results which were compatible with both the DGB motif theory and the charge dependency rule (Rizzo et al., 2019). However in the present thesis, many tails were shown to successfully bind to GOLPH3 in vitro which satisfy the charge dependency rule but do not bear a DGB such as the tails for GALNT2 (MRRRSR), MANEAL (MARRRRRA) and CHSY1 (MARRGRR) as a small selection of examples (Figure 4.6). Future structural insights are required to clarify the GOLPH3 recognition motif and to map the tail-binding site on GOLPH3. Furthermore, bioinformatic analysis of the cytoplasmic tails in the GOLPH3+3L interactome may yield some insights into the motif.

One issue to reconcile is the GOLPH3 charge dependency rule in the context of topogenesis. According to the positive-inside rule, to aid the insertion of proteins into the membrane in the correct topology at the ER, proteins exhibit an enrichment of membrane-proximal positively-charged residues in their cytosolic tails (Baker et al., 2017; von Heijne, 1986). This includes not just integral membrane proteins of the early secretory pathway but also those destined for post-Golgi locations. This leads to the question: if integral membrane proteins of post-Golgi compartments have an enrichment of positively charged residues in their cytoplasmic tails, then how do they evade the promiscuous charge-dependent filter that is GOLPH3 at the trans-Golgi? One answer is that despite having enough membrane-proximal positive charges for correct topogenesis, they simply have enough membrane-distal negative charges to dissatisfy the GOLPH3 charge dependency without upsetting membrane insertion. An enrichment of negatively-charged residues on the endoplasmic side could also aid in the fidelity of insertion according to the negative-outside adjunct to the topogenesis rule and this would not confer GOLPH3 recognition (Baker et al., 2017). Alternatively, post-translational modification (PTM) could be leveraged

as a means to evade GOLPH3-dependent Golgi retention. A good illustration of that is sucrase-isomaltase, its cytoplasmic tail is phosphorylated which effectively adds a negative charge but only after it has been correctly inserted into the membrane (Keller et al., 1995). When the predicted phosphoserine and another negatively-charged residue were simultaneously targeted by mutagenesis, the charge threshold was satisfied and GOLPH3 could bind to the tail (Figure 4.7A). Phosphorylation may not be the only PTM utilised for GOLPH3 evasion. A recent study has suggested that the S-palmitoylation of cytoplasmic tails acts as a cue for the post-Golgi sorting of anterograde cargo (Ernst et al., 2018). It is plausible that the S-palmitoylation, and other lipidation modifications, could serve as evasion signals by anchoring the cytoplasmic tails to the membrane thus rendering the positive charges inaccessible to GOLPH3.

While the current study has focused exclusively on type II transmembrane cargo, a smaller proportion of the proteins of the GOLPH3+3L interactome were of different topologies such as type I and multi-pass (Table 4.2.1, Supplementary Table 8.5). It remains to be seen if the charge dependency rule is consistent with GOLPH3 clients of these varying topologies. In this context, it is important to consider that the amino terminus of a protein is positively-charged and the carboxy terminus negatively-charged. On this basis, it is likely that GOLPH3+3L cargo selection is biased towards the cytoplasmic N-termini of type II cargo over the cytoplasmic C-termini of type I cargo. Some GOLPH3+3L interactors such as giantin and golgin-84 have cytoplasmic-facing N-terminal sections which are vastly extended in relation to many of the glycan-modifying enzyme clients investigated. Based on the small size of the enzyme tails and that the fact that GOLPH3 interacts directly with the bilayer, it is highly likely that GOLPH3 charge-recognition of the tails has a positional-dependency. Moreover, based on the size of the peptide tails shown to bind in vitro, it is likely they are too short to exhibit any secondary structure. Extrapolation from the binding results of these short-tailed clients would suggest that if GOLPH3+3L are interacting with proteins such as giantin and golgin-84, it is likely through membrane-proximal stretches which are disordered and positively-charged. Initial attempts were made to characterise positional-dependency of GOLPH3-binding in vitro (Figure 4.7B); however, more carefully designed future studies are required.

In summary, FAM114A2 was shown to bind *in vitro* to the cytoplasmic tails of a host of glycan-modifying enzymes through a membrane-proximal polybasic motif. For GOLPH3 and GOLPH3L, potentially over a hundred novel clients were identified suggesting GOLPH3 proteins function as master COPI adaptors for intra-Golgi vesicles at the trans-Golgi. For GOLPH3 recognition, the cytoplasmic tails appeared to require a total charge of ≥4+. To interrogate the different cytoplasmic tail recognition motifs in a cellular context and to understand the function of the FAM114A family members, the next chapter will discuss numerous cell-based assays and analyses conducted with GOLPH3 and FAM114A CRISPR knockout cell lines.

5. *In vivo* analysis of FAM114A and GOLPH3 family proteins

5.1 Introduction

In the previous chapter, FAM114A2 was shown to interact with the cytoplasmic tails of a subset of COPI cargo *in vitro*. Moreover, GOLPH3+3L binding studies potentially identified in excess of a hundred novel clients. In order to investigate the cellular relevance of these *in vitro* interactions and to explore the functions of both families of proteins, this chapter will outline a number of phenotypic analyses of FAM114A and GOLPH3 CRISPR-Cas9 deletion mutants. FAM114A2 was shown to interact with cytoplasmic tails through a membrane-proximal polybasic motif while GOLPH3 interacted with tails with a total charge of ≥3+. In this chapter these cytoplasmic tail recognition motifs will be validated in the context of the cell.

5.2 CRISPR-Cas9 gene editing of FAM114A and GOLPH3 family members

In order to explore the functional role of FAM114A family proteins and their relationship to the GOLPH3 family, a series of combinatorial FAM114A and GOLPH3 knockout U2OS cells lines were generated by CRISPR-Cas9 gene editing (Figure 5.1A). Of the various knockouts; *ΔΔFAM114A1, FAM114A2* cells showed no clear defects in Golgi morphology, polarity or a depletion of any of the markers relative to wild-type U2OS cells (Figure 5.1B). This was in contrast to *ΔΔGOLPH3, GOLPH3L* and *ΔΔΔΔFAM114A1, FAM114A2, GOLPH3, GOLPH3L* cells which displayed a marked reduction in the signal from the COPI cargo markers GALNT2 and giantin relative to wild-type cells. There was no clear additive effect when comparing the *ΔΔGOLPH3, GOLPH3L* cells to the quadruple knockout. The cis-Golgi marker GM130 showed no clear change in signal intensity or morphology between the genotypes suggesting the defects may be specific to COPI cargo as opposed to the Golgi en masse. Whilst maintaining cells in culture; *ΔΔGOLPH3, GOLPH3* and quadruple knockout cells displayed a notable defect in growth rate relative to wild-type U2OS cells (data not shown). In contrast, FAM114A single and double knockout cells grew at a comparable rate to wild-type cells. In summary, the deletion of GOLPH3 family genes, but not FAM114A family genes, caused the depletion of COPI cargo and a subtle growth defect.

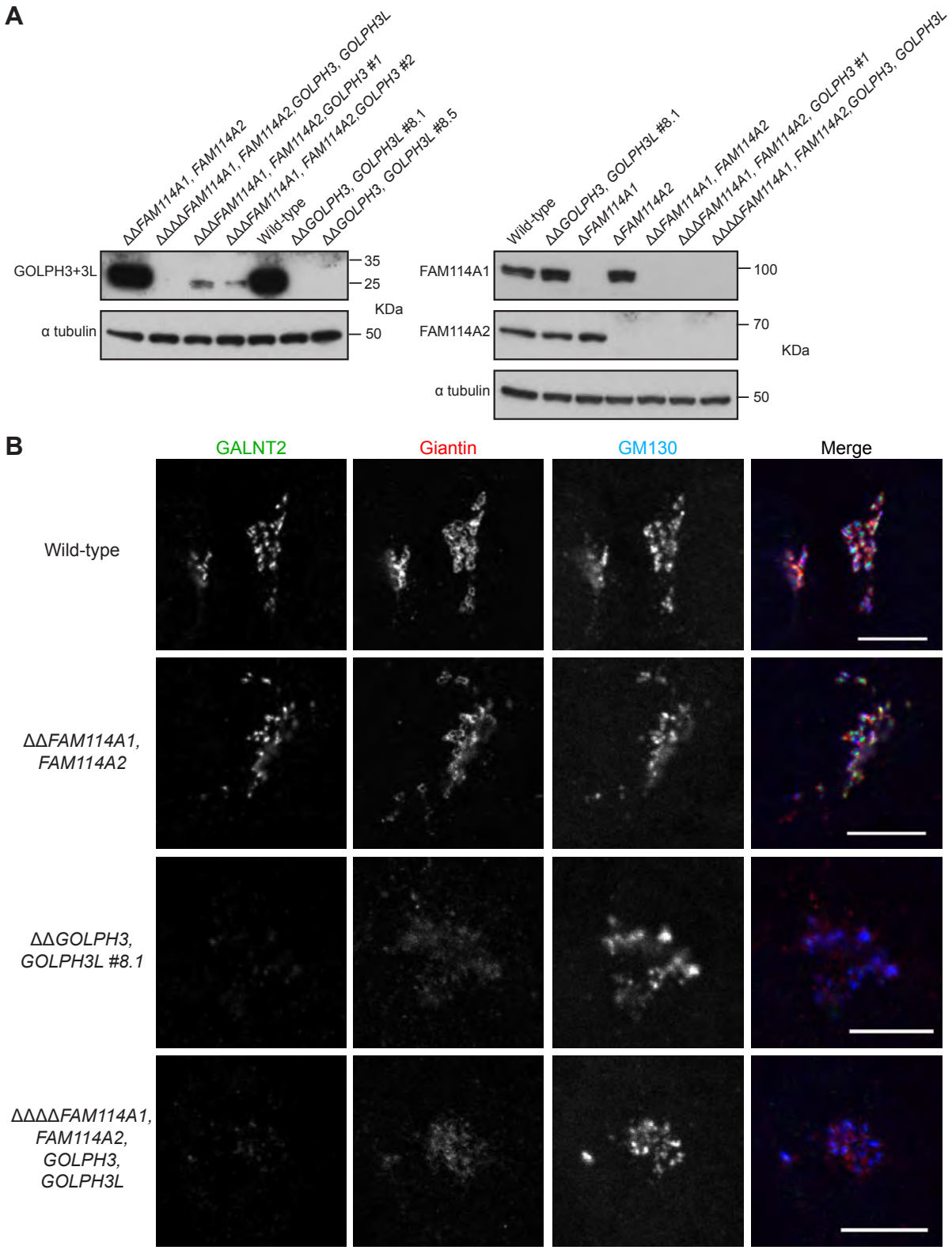


Figure 5.1 Deletion of GOLPH3 family but not FAM114A family genes causes a depletion in COPI vesicle cargo. (**A**) Immunoblots of FAM114A and GOLPH3 family combinatorial U2OS CRISPR knockout cell lines. (**B**) Micrographs displaying different U2OS CRISPR knockout cell lines. Cells are stained for GALNT2 and giantin as different COPI vesicle cargo markers and GM130 as a cis-Golgi marker. Images were taken using identical laser voltages and modified with identical contrast settings. Scale bars correspond to 10 μm.

5.3 A flow cytometry-based *in vivo* Golgi retention assay

Numerous reports have implicated the cytoplasmic tail and the TMD of glycan-modifying enzymes in Golgi-targeting via distinct mechanisms (Munro, 1991; Schmitz et al., 2008; Tu et al., 2008). In order to explore these mechanisms, to conduct further phenotypic analyses on the knockouts and to examine the proposed GOLPH3 consensus motif in a cellular context; a flow-cytometry based *in vivo* Golgi retention assay was developed (Figure 5.2A). A series of chimeric reporters were designed in which the cytoplasmic tail or TMD of the plasma membrane reporter sucrase-isomaltase was substituted for that of different Golgi enzymes and the chimeras fused to a GFP-FLAG tag. When expressed in cells, reporters which are subject to mechanisms of Golgi retention will be largely retained in intracellular Golgi-associated membrane compartments. Therefore, the GFP tag will not accessible to an anti-GFP Alexa Fluor 647 (A647) antibody stain under non-permeabilising conditions. In contrast, reporters which are not subject to Golgi retention should be able to passage through to the plasma membrane. Once at the cell surface, the GFP tag becomes accessible to the anti-GFP A647 antibody. By taking a ratio of the A647 signal (which equates to the cell surface signal) and the GFP signal (which equates to a total cell signal), it is possible to derive a quantitative readout for Golgi retention. This assay can dissect out the Golgi retention cues in the Golgi enzyme tails or TMDs and can be combined with genetics tools to test the factors responsible for mediating this retention (e.g. COPI adaptors).

The reporters were integrated into the various GOLPH3 and FAM114A family knockout cell lines using the PiggyBac transposase system. Cells were subject to drug selection to derive polyclonal cell lines. Polyclonal cell lines were favoured over monoclonal cell lines as it was considered beneficial to have a range of expression levels so as to gauge the saturation points of the different retention mechanisms. As the reporters were integrated under a cumate-inducible promoter, cells were cultured in the presence of cumate prior to analysis. As an initial proof of concept, the reporter cell lines were counterstained for the COPI vesicle and Golgi marker golgin-84 in addition to a GFP-booster prior immunofluorescence imaging. The sucrase-isomaltase plasma membrane reporter displayed a robust cell surface localisation in both wild-type and *ΔΔGOLPH3, GOLPH3L* U2OS cells (Figure 5.2B). ST6GAL1 was chosen as a model Golgi enzyme as its relatively short TMD has been previously shown to be sufficient for Golgi targeting (Munro, 1991). Moreover, the tail of ST6GAL1 has been shown to interact with GOLPH3 to facilitate incorporation of the enzyme into COPI vesicles *in vitro* (Eckert et al., 2014). In line with previous reports, the ST6GAL1 TMD chimera showed a robust Golgi localisation with considerable overlap with

golgin-84 in both wild-type and *ΔΔGOLPH3, GOLPH3L* U2OS cells. This confirmed that the reporter localised to the Golgi as opposed to other intracellular membranes and that the TMD of ST6GAL1 was sufficient to confer Golgi retention in a GOLPH3+3L-independent manner. In contrast, the ST6GAL1 cytoplasmic tail chimera exhibited a strong Golgi localisation in wild-type cells but in *ΔΔGOLPH3, GOLPH3L* U2OS cells a considerable proportion localised at the plasma membrane in addition to the Golgi. This supported previous findings in showing that the tail of ST6GAL1 can retain the enzyme at the Golgi in a GOLPH3+3L-dependent manner. Next a GALNT2 cytoplasmic tail chimera was chosen for analysis as it bound *in vitro* to both FAM114A2 and GOLPH3 (Chapter 4), a previously unreported interaction in both instances. As with the ST6GAL1 cytoplasmic tail chimera, the GALNT2 cytoplasmic tail chimera localised to the Golgi in wild-type cells but exhibited a dual plasma membrane/Golgi localisation in *ΔΔGOLPH3, GOLPH3L* cells. In contrast, dual deletion of *FAM114A1* and *FAM114A2* had no effect upon the localisation of the GALNT2 cytoplasmic tail chimera and there was no clear additive effect observed in the quadruple knockout cell line in relation to *ΔΔGOLPH3, GOLPH3L* cell line. In summary, despite the cytoplasmic tail of GALNT2 interacting with both FAM114A2 and GOLPH3 *in vitro*, only GOLPH3 was required for tail-dependent Golgi-targeting *in vivo*.

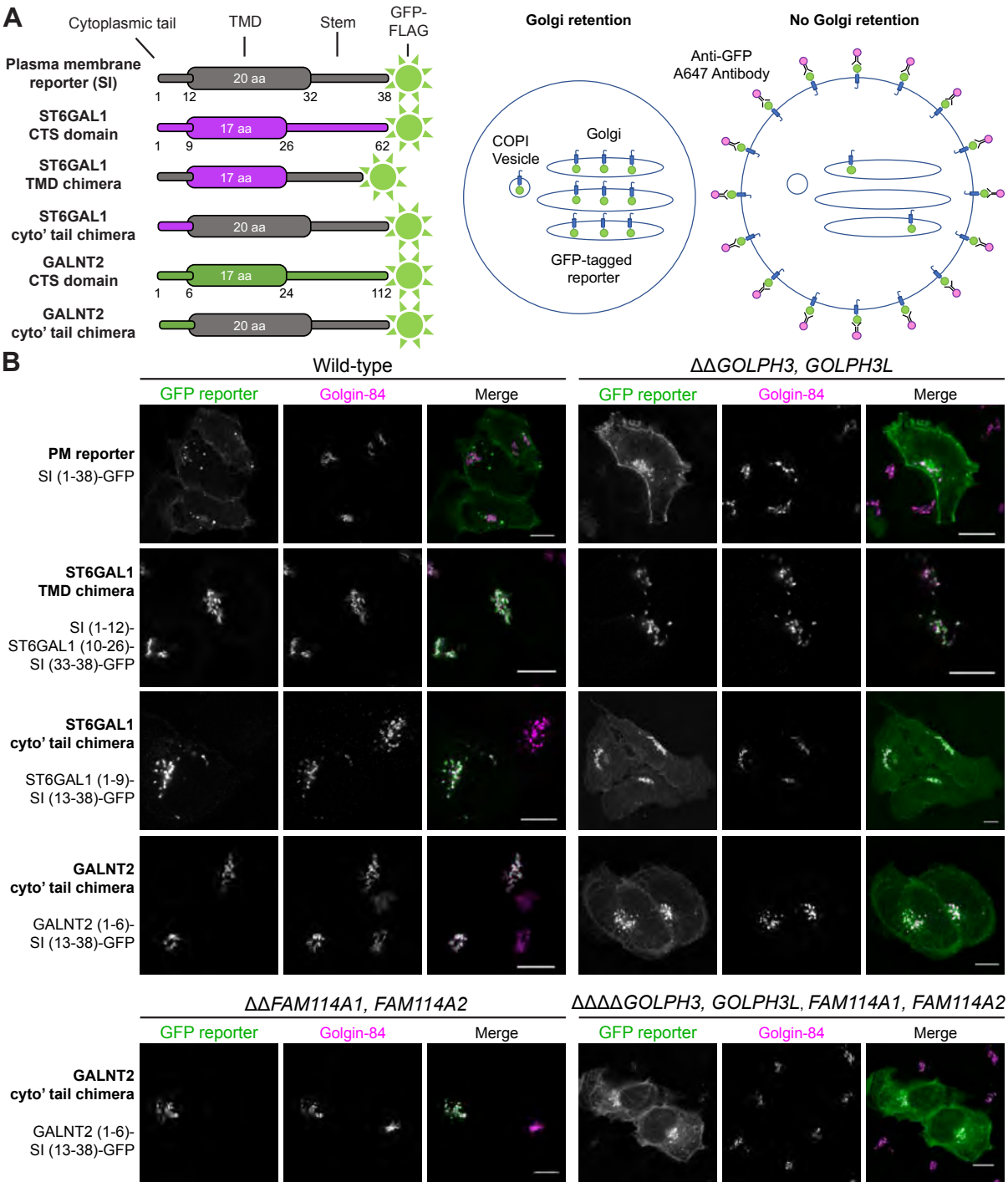


Figure 5.2 An *in vivo* Golgi retention assay to interrogate cytoplasmic and TMD Golgi sorting cues and the Golgi retention capacity of GOLPH3 and FAM114A family proteins. (**A**) A cartoon schematic showing the basis of the *in vivo* Golgi retention assay. Different chimeric GFP-tagged type II transmembrane reporters were generated in which the cytoplasmic tail or TMD of different Golgi enzymes were substituted for that of the plasma membrane reporter sucrase-isomaltase. Under conditions of Golgi retention, the reporter is sequestered in intracellular compartments (primarily the Golgi and COPI vesicles) and therefore the lumenal GFP-FLAG tag is inaccessible to antibodies under non-permeabilising conditions. In contrast, reporters which are not subject to Golgi retention can passage through to the plasma membrane where the GFP-FLAG tag becomes extracellular and therefore accessible to an anti-GFP Alexa Fluor 647 (A647)-conjugated antibody. Antibody-stained cells are analysed by flow cytometry and a ratio of the A647 and GFP signals is used as a read out for retention. (**B**) Proof-of-principle micrographs displaying stable cells lines expressing GFP-tagged reporters in different U2OS CRISPR knockout cell lines. Cells are counterstained for golgin-84 as a COPI and Golgi marker and with a GFP booster. Scale bars correspond to 10 μm.

Since the proof of principle was successful, the assay was expanded to look at a wide array of different reporters and the mode of analysis was switched from immunofluorescence to flow cytometry. Across all of the U2OS genetic backgrounds tested, the plasma membrane reporter exhibited a high cell surface signal to total cell signal ratio indicative of minimal Golgi retention (Figure 5.3). In line with the immunofluorescence images, the GALNT2 cytoplasmic tail chimera displayed robust Golgi retention in wild-type cells which was mostly relieved in *ΔΔGOLPH3, GOLPH3L* and quadruple knockout cells but not in *ΔΔFAM114A1, FAM114A2* cells. Moreover, there appeared to be no additive effect when comparing the *ΔΔGOLPH3, GOLPH3L* cell line with the quadruple knockout. The ST6GAL1 cytoplasmic tail chimera also exhibited GOLPH3+3L-dependent Golgi retention but the retention was less robust than observed for the analogous GALNT2 chimera. The enzymes also displayed a discrepancy in their TMD-dependent retention mechanisms. While the ST6GAL1 TMD chimera was retained at the Golgi in a GOLPH3+3L-independent manner, the GALNT2 TMD chimera was not retained in wild-type U2OS cells. Previous models have proposed that TMD-dependent sorting is linked to the physiochemical properties of the relatively short TMDs (Sharpe et al., 2010, Chapter 1); however, both TMDs are predicted to be identically short (ST6GAL1 [FSCCVLVFLLFAVICVW], GALNT2 [MLLCFAFLWVLGIAYYMY] - 17 amino acids; sucrase-isomaltase (LIVLFVIVTIIAIALIVVLA) – 20 amino acids; Figure 5.2A). More TMDs will need to be investigated before firm conclusions about the membrane thickness model can be drawn from this assay.

GOLPH3 was shown to interact with the cytoplasmic tails of type II transmembrane proteins with a total predicted charge of 4+ or higher (Figure 4.6, 4.7). Binding experiments showed that the membrane-proximal insertion of 3 positively-charged residues into the tail of sucrase-isomaltase was sufficient to satisfy the charge threshold and sufficient to confer GOLPH3-binding *in vitro* (Figure 4.7B). To see if this was also sufficient to relocate sucrase-isomaltase from the plasma membrane to the Golgi, some of these tails were tested in the *in vivo* Golgi retention assay. In accordance with the *in vitro* findings, a triple arginine or lysine insertion into the tail of the plasma membrane reporter was sufficient to confer Golgi retention in a GOLPH3+3L-dependent manner. A triple histidine insertion into the tail of the plasma membrane reporter was not sufficient to meet the charge threshold and the chimera did not bind to GOLPH3 *in vitro* (Figure 4.7B). Curiously, the triple histidine insertion resulted in the robust Golgi retention of the reporter which was only slightly perturbed in the *ΔΔGOLPH3, GOLPH3L* background (Figure 5.3). It was confirmed by immunofluorescence that the reporter localised to the Golgi and not to other intracellular compartments (data not shown). It appears a triple histidine insertion into the tail of sucrase-isomaltase, initially intended to act as a negative control, may have inadvertently unearthed a novel GOLPH3-independent Golgi retention mechanism. The TPST1

cytoplasmic tail chimera was also included in the assay as it failed to meet the GOLPH3 charge threshold and failed to bind to GOLPH3 *in vitro*. The tail of TPST1 (MVGKLKQN) was sufficient to retain the reporter at the Golgi in the *in vivo* Golgi retention assay but in a GOLPH3-dependent manner (Figure 5.3). The discrepancy between a lack of *in vitro* GOLPH3-binding and *in vivo* GOLPH3-dependent Golgi retention remains the focus of future experimentation.

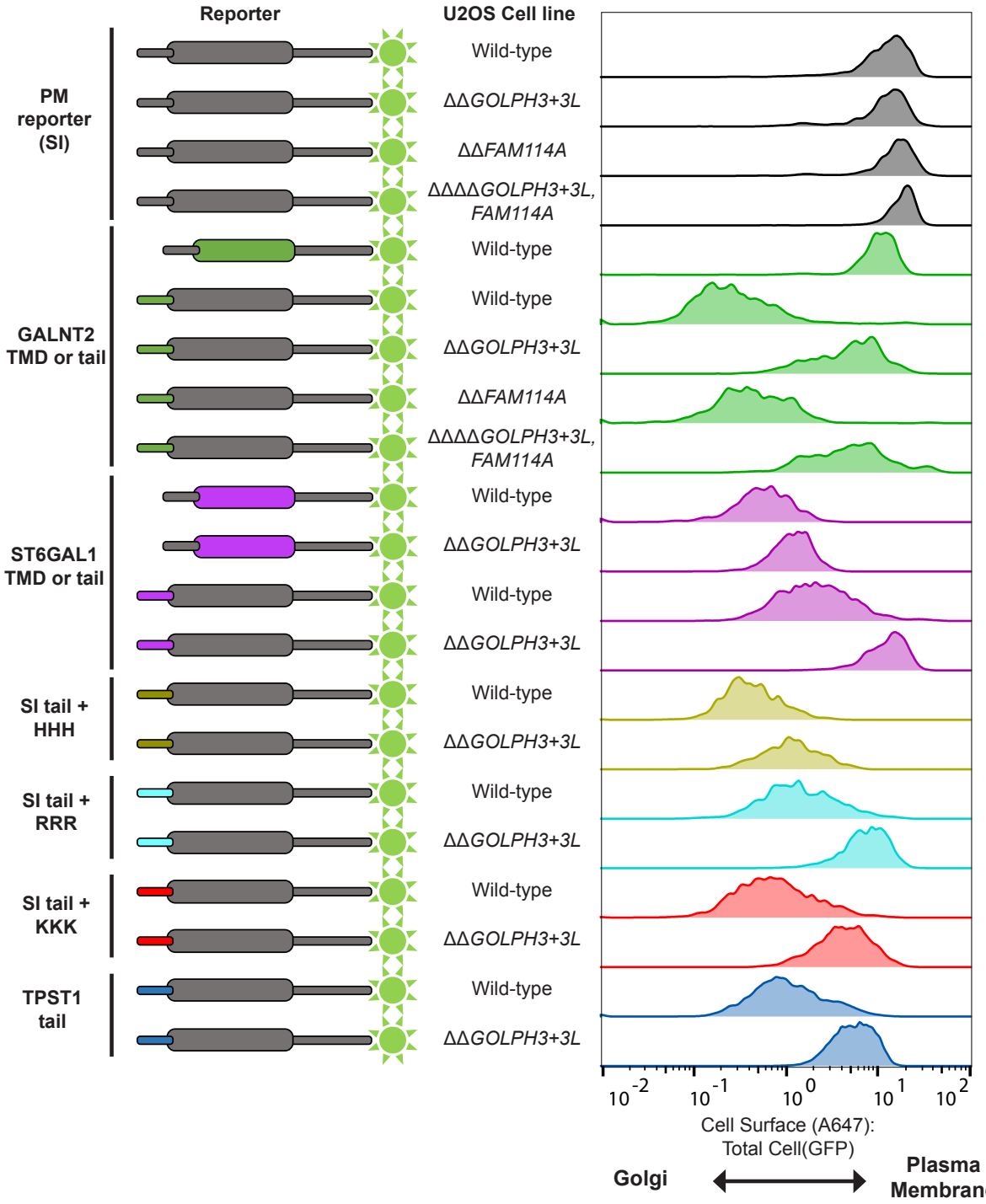


Figure 5.3 The cytoplasmic tail and TMD of different Golgi enzymes can bestow Golgi retention via GOLPH3+3L-dependent and independent mechanisms. Shown are the results from the flow cytometry-based *in vivo* Golgi retention assay to investigate Golgi retention mechanisms. Histograms (see right) were generated from analysis of polyclonal cumate-inducible stable cells lines expressing different chimeric reporters (see left) made from U2OS cell lines of different CRISPR knockout backgrounds (see middle). Cells were cultured in the presence of cumate for at least a week prior to analysis. Histograms are normalised to the mode value and display ratios of the cell surface signal to total cell signal serving as a readout for Golgi retention. At least 10,000 events were collected for each cell line. Singlets were gated for based on forward and side scatter, dead cells were excluded using a fixable live-dead stain. Shown are the results of a single repeat of a triplicate.

5.4 Upon the deletion of GOLPH3 genes, clients are targeted to the lysosome for degradation

In the reports which initially characterised the *Vps74* knockout yeast strain, it was demonstrated that cargo was mislocalised to the vacuole for degradation (Schmitz et al., 2008; Tu et al., 2008). Similarly, simultaneous deletion of the mammalian orthologs *GOLPH3* and *GOLPH3L* in U2OS cells caused a marked depletion of cargo clients GALNT2 and giantin. In order to validate the lysosomal mislocalisation and degradation of cargo, wild-type and *ΔΔGOLPH3, GOLPH3L* U2OS cell lines were treated with genetic and pharmacological rescues. Levels of the COPI cargo protein GALNT4 were tracked by western blot as a marker for degradation. Cells were transiently transfected with a construct encoding BFP and a P2A-T2A tandem self-cleaving peptide (Liu et al., 2017) upstream of GOLPH3 or a mutant GOLPH3 with two conserved arginines mutated (GOLPH3 RRAA) which is reported to ablate binding to coatomer (Tu et al., 2012). After 24 hours of overexpression of the GOLPH3 fusions, there was a slight recovery in GALNT4 levels in the double knockout cells when compared to untransfected cells (Figure 5.4A). Reasons for the incomplete recovery are that it is likely a proportion of the cells were unsuccessfully transfected and cells were harvested too soon after transfection. There was no difference in rescue between the wild-type GOLPH3 and GOLPH3 RRAA treatment suggesting the capacity for rescue may be independent of coatomer binding. In order to test the lysosomal dependency of the phenotype, cells were treated with the lysosomal inhibitor bafilomycin and the proteasomal inhibitor MG132 for 24 hours and 4 hours respectively. Bafilomycin treatment was sufficient to rescue GALNT4 to levels comparable to wild-type cells but there was no observed change after MG132 treatment. This confirms that the depletion of cargo occurs specifically via a lysosomal pathway. Considering previous reports, the finding that GOLPH3 RRAA could rescue comparably to wild-type GOLPH3 was surprising (Tu et al., 2012). To biochemically validate the interaction between GOLPH3 and coatomer, a pair of mutants were designed for expression in bacteria which were expected to ablate binding to coatomer. GST-tagged wild-type GOLPH3, an N-terminal GOLPH3 deletion mutant (Δ1-51) and the RRAA mutant were purified from bacteria and used in a GST pulldown to test if they could pull-out coatomer from HEK293T cell lysate. Since coatomer forms a stable complex *in vitro*, it was sufficient to blot for the β-COP subunit as a marker for the complex. When variation in bait expression was considered, both mutants bound to coatomer as efficiently as the wild-type protein while the GST negative control displayed only very weak non-specific binding (Figure 5.4B). This may explain why the capacity of the GOLPH3 RRAA mutant to rescue is comparable to that of the wild-type protein.

The mechanism by which the cargo is targeted to the lysosome from the Golgi in *ΔΔGOLPH3, GOLPH3L* knockout cells is not clear. In order to dissect out which features of the cargo serve as a cue for degradation in a knockout background and to identify the proteins involved in this pathway, a flow cytometry-based lysosomal degradation assay was designed. The CTS domain of GALNT2 was fused to a GFP tag and was downstream of a tandem P2A-T2A self-cleaving peptide and mCherry (Figure 5.4C). The reporter was then stably integrated into wild-type or *ΔΔGOLPH3, GOLPH3L* U2OS cells. The expectation of the assay was that in a knockout background, the GFP-tagged GALNT2 reporter would be degraded more readily while the rate of degradation of the mCherry molecule would remain constant. Therefore, by taking a ratio of the GFP signal to the mCherry signal, it would be possible to derive a robust quantitative readout for the stability of the GALNT2 reporter. When these reporter cell lines were analysed by flow cytometry, across the population there was a shift in the GFP to mCherry ratio indicative of a reduced stability of the GALNT2 reporter in the double knockout cell line compared to the wild-type background (Figure 5.4C). Future experiments with different chimeric reporters will serve to interrogate the specific cues which trigger degradation in the double knockout background. Moreover, the flow cytometry-based nature of the assay lends itself to future CRISPR-Cas9 genetic screens for novel factors associated with this degradative pathway.

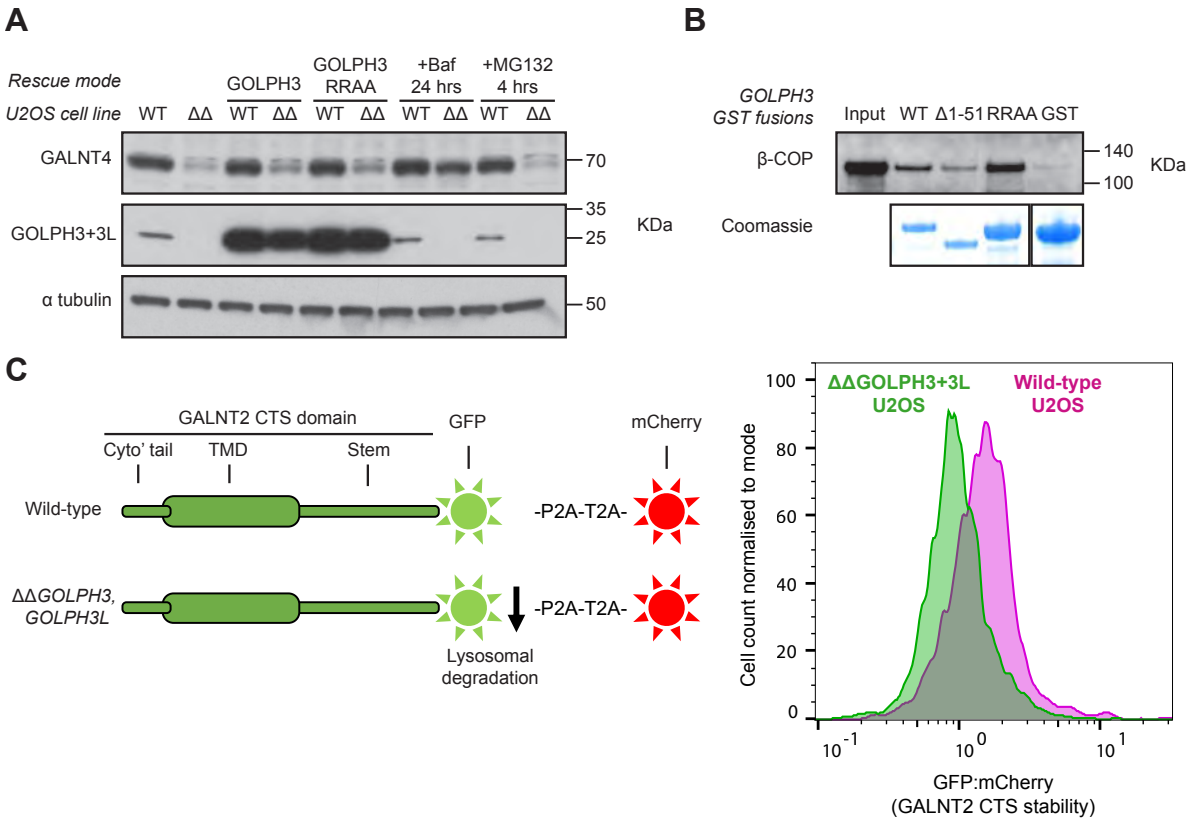


Figure 5.4 COPI cargo is targeted to the lysosome for degradation in *ΔΔGOLPH3, GOLPH3L* U2OS cells. (**A**) Immunoblots displaying genetic and pharmacological rescue of COPI cargo protein GALNT4. Cells were harvested 24 hours after transfection with a construct encoding BFP-P2A-T2A-wild-type GOLPH3 or an RR14,15AA point mutant GOLPH3 which was reported to ablate the interaction with the COPI coat (Tu et al., 2012). Cells were alternatively treated with a lysosomal inhibitor (bafilomycin, 160 nM) or proteosomal inhibitor (MG132, 10 μM). N=1. (**B**) An immunoblot and Coomassie stain from a GST pulldown experiment testing the ability of GST-tagged GOLPH3 mutants to interact with the COPI coat from HEK293T cell lysate. N=1. (**C**) A schematic outlining a flow cytometry-based assay to acquire a quantitative readout for lysosomal degradation. A reporter consisting of the CTS domain of GALNT2 fused to GFP, a tandem self-cleaving P2A-T2A peptide sequence followed by mCherry was stably and randomly integrated into U2OS cells of a wild-type or *ΔΔGOLPH3, GOLPH3L* U2OS background. Under conditions of increased lysosomal degradation of the GALNT2 reporter, the GFP signal is expected to decrease while the soluble mCherry signal would remain constant. Shown are histograms of the ratio of the GFP to mCherry signal which serve as a readout for the stability of the reporter/rate of lysosomal degradation. Histograms are normalised to the mode value. At least 10,000 events were collected for each cell line. Singlets were gated for based on forward and side scatter, dead cells were excluded using a fixable live-dead stain.

5.5 Deletion of GOLPH3 genes causes a global defect in glycosylation

Since many glycan-modifying enzyme clients are mislocalised and degraded in *ΔΔGOLPH3, GOLPH3L* knockout cells, it would be expected that there would be downstream defects in the glycosylation status of the cell. Furthermore, since FAM114A2 interacts with glycan-modifying enzymes, investigation of the glycosylation status of FAM114A knockout cell lines may help to illuminate the elusive phenotype and function of this family of proteins. To explore this, U2OS cells of the different GOLPH3 and FAM114A knockout backgrounds were stained with a panel of fluorescein-conjugated lectins with a range of different sugar specificities. A negative control sample was included to validate the specificity of each lectin in which cells were incubated with the lectin and an excessive concentration of a competitive sugar. For all the lectins tested, there were considerable shifts in lectin staining intensity in *ΔΔGOLPH3, GOLPH3L* and quadruple knockout cells when compared to wild-type cells (Figure 5.5). Furthermore, there were no additive effects observed when comparisons were made between the *ΔΔGOLPH3, GOLPH3L* cell line and the quadruple knockout cell line. No effect or only subtle effects were observed for single and double FAM114A knockout cell lines in relation to the wild-type cell line.

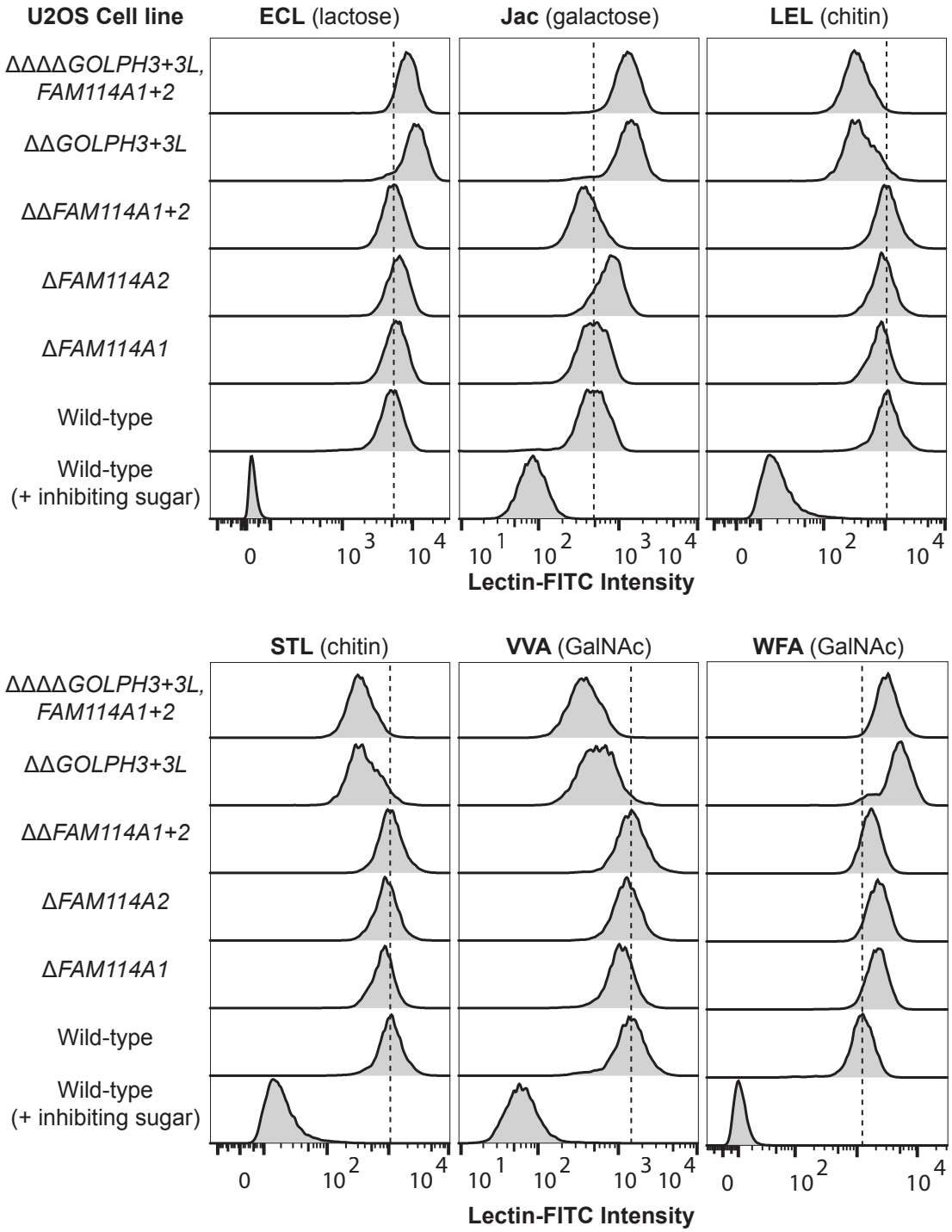


Figure 5.5 Simultaneous deletion of *GOLPH3* and *GOLPH3L* in U2OS cells causes global defects in glycosylation while deletion of FAM114A genes have mild effects or no effect on glycosylation. Histograms generated from flow cytometry analysis of different CRISPR knockout U2OS cells line subjected to cell surface stains with a panel of different FITC-conjugated lectins (see top, in bold). Specificity of the lectin was validated in which cells were stained in the presence of saturating concentrations of a competing sugar (see top, brackets). Histograms are normalised to the mode value for each treatment. Dotted lines mark the mode intensity value for wild-type cells. At least 10,000 events were collected for each cell line. Singlets were gated for based on forward and side scatter, dead cells were excluded using a fixable live-dead stain. Shown are the results of a single repeat of a triplicate.

5.6 Discussion

Work by Liu and colleagues had demonstrated that the substitution of the cytoplasmic tail of sucrase-isomaltase for that of GALNT2 was sufficient to target the reporter to the Golgi; however, they did not describe the mechanism by which this occurred (Liu et al., 2018). By utilising the same reporter in the *in vivo* Golgi retention assay in this thesis, the cytoplasmic tail of GALNT2 and ST6GAL1 were shown to be sufficient for Golgi retention in a GOLPH3+3L-dependent manner. The assay also validated the charge dependency rule for GOLPH3 recognition determined from *in vitro* binding studies (Chapter 4). Sucrase-isomaltase tail mutants with a charge of ≥4+ were retained at the Golgi in a GOLPH3+3L-dependent manner. A surprising result was that the TPST1 cytoplasmic tail chimera, which failed to bind to GOLPH3 *in vitro* (Figure 4.7A), was retained at the Golgi in a GOLPH3+3L-dependent manner. TPST1 has been reported to form oligomers with ST6GAL1 so it may be possible that TPST1 is retained by GOLPH3 through an indirect interaction as part of a heteromeric with endogenous ST6GAL1 (Hartmann-Fatu et al., 2015). However, only the cytoplasmic tail of TPST1 is present in the chimera and therefore this explanation requires for the short tail (only 8 amino acids) to be sufficient to form oligomers with ST6GAL1, a possibility that seems unlikely. Another curious result was that a membrane-proximal triple histidine insertion into the cytoplasmic tail of sucrase-isomaltase was sufficient for robust Golgi retention. Moreover, this mutant does not bind to GOLPH3 *in vitro* and the retention *in vivo* is GOLPH3+3L-independent. Could this represent a serendipitous discovery of a novel Golgi retention motif in cytoplasmic tails? If so, is there a novel COPI adaptor or Golgi retention factor involved in this mechanism? Since the *in vivo* Golgi retention assay has been conceptually validated and is flow cytometry-based with a robust read-out, it is perfectly set-up for future CRISPR-Cas9 genetic screens to identify novel Golgi retention factors.

The *in vivo* Golgi retention assay also showed that the TMD of ST6GAL1 contained a Golgi retention signal, a finding in accordance with reports from nearly 30 years ago (Munro, 1991). Moreover, the retention was GOLPH3+3L-independent, which is consistent with the idea that GOLPH3 proteins specifically recognise cytoplasmic tails. The membrane thickness model posits that Golgi enzymes are retained at the Golgi based on the physiochemical properties of their relatively short TMDs (Bretscher and Munro, 1993; Sharpe et al., 2010; Chapter 1). The Golgi enzymes favour residence in thinner sphingolipid/sterol-poor membranes of the ER/early-Golgi as opposed to thicker sphingolipid/sterol-rich membranes of the trans-Golgi and post-Golgi compartments (Bretscher and Munro, 1993; Sharpe et al., 2010; Chapter 1). ST6GAL1 has a short TMD of

17 amino acids and therefore the finding that the TMD is sufficient for Golgi retention is consistent with this hypothesis. However, the TMD of GALNT2, which is also predicted to be 17 amino acids long, was insufficient to impart Golgi retention. Since only two TMDs have been tested in this assay it is difficult to draw conclusions pertaining to the membrane-thickness model. It could be possible that there are subtle nuances associated with the physiochemical nature of a TMD which are required to complement their short length to impart Golgi retention. An alternative model is that the TMD of ST6GAL1 could contain a sequence-specific retention motif which is recognised by a membrane-spanning COPI adaptor. Recently, a sequence-specific Golgi retention motif was characterised in the TMD of a plant Golgi enzyme (Schoberer et al., 2019). Of topical relevance, Orf7b of the severe acute respiratory syndrome coronavirus (SARS-CoV) family was shown to have a sequence-specific Golgi-targeting signal in its TMD (Schaecher et al., 2008). The TMD of Orf7b shares a striking sequence similarity with ST6GAL1 (ST6GAL1 – FSCCVLVFLLFAVICVW, Orf7b – FYLCFLAFLLFLVLIMLIIFWF; 45.4% sequence identity) and it is therefore plausible they are targeted to the Golgi by the same machinery. The TM9SF family, also known as nonaspanins, are an evolutionary-conserved family of Golgi-resident multi-pass membrane proteins which have been shown to interact with both the COPI coat and the TMD of a Golgi enzyme (Woo et al., 2015; Yamaji et al., 2019). They have also appeared as candidates in numerous CRISPR-Cas9 screens centred on readouts linked to the glycosylation status of a cell (Pacheco et al., 2018; Tanaka et al., 2017; Tian et al., 2018; Yamaji et al., 2019). For these reasons, they are prime suspects in the hunt for a potential TMD-associating COPI adaptor. As alluded to earlier, the *in vivo* Golgi retention assay is perfectly poised for interrogating the TM9SF family as novel Golgi retention factors using CRISPR-Cas9 gene editing.

Analysis of an extensive data set shows that ER and Golgi-residents have shorter TMDs than their post-Golgi counterparts across eukaryotic life (Sharpe et al., 2010). Furthermore, there is compelling evidence to show that the membrane increases in thickness between the ER and cis/medial-Golgi compared to the trans-Golgi and post-Golgi compartments (Bykov et al., 2017). If the short length of the TMDs is not responsible for Golgi retention *per se*, then what is the functional significance of this feature? The deletion of *GOLPH3* and *GOLPH3L* triggered the lysosomal degradation of COPI cargo. It is plausible that by eliminating the GOLPH3+3L-dependent recycling of enzymes to earlier cisternae, the enzymes will begin to accumulate in the incompatibly-thick membranes of the late-Golgi. Of course, this would not be the case for enzymes which are retrieved in a GOLPH3+3L-independent manner. The ensuing hydrophobic mismatch of enzymes mislocalised in the late-Golgi could then somehow serve as a cue for lysosomal degradation. In other words,

instead of the short TMDs serving retention directly, they could play a homeostatic role in acting as a signal for the turnover of enzymes incorrectly localised at the late-Golgi or post-Golgi compartments. This may be an important mechanism to control the distribution of glycan-modifying enzymes and therefore ensure the fidelity of cellular glycosylation.

The degradation of clients in *Vps74/GOLPH3* knockout cells is known to be via a vacuolar/lysosomal pathway; however, the details of this pathway remain unclear. A lysosomal degradation assay was developed to explore this pathway which has been partially conceptually validated. This assay will be used to confirm if a short TMD, or another feature of a mislocalised client, is the cue for lysosomal targeting and degradation. As with the *in vivo* Golgi retention assay, the lysosomal degradation assay was designed for flow-cytometry analysis so it would generate a quantitative readout which is sensitive to subtle variations and so it was tractable for CRISPR-Cas9 genetic screening. This could be leveraged to identify the factors responsible for targeting the clients to the lysosome. Interesting *Vps74* knockout yeast strains display a negative genetic interaction with the Golgi-localized, γ-ear-containing, ADP ribosylation factor (ARF)-binding proteins (GGAs), GGA1 and GGA2 (Wood et al., 2012). This family of proteins function as clathrin adaptors in trans-Golgi to endolysosomal trafficking pathways (Uemura and Waguri, 2020). It is plausible that when both *Vps74* and the GGAs are deleted, the mislocalised Vps74 clients can no longer be turned over efficiently leading to prolonged hydrophobic mismatch and protein aggregation. The formation of protein aggregates and/or aberrant glycosylation resulting from mislocalised Golgi enzymes could result in a loss of fitness that is responsible for this negative genetic interaction. It will be intriguing to see if genetic ablation of the GGAs, or pharmacological inhibition of the lysosome, can rescue the phenotype seen in *ΔΔGOLPH3, GOLPH3L* cells in the lysosomal degradation assay.

Curiously, GALNT4 was targeted for lysosomal degradation in *ΔΔGOLPH3, GOLPH3L* cells despite *in vitro* binding studies suggesting it was not a direct GOLPH3 client (Figure 4.7A). It is plausible that GALNT4 could indirectly associate with GOLPH3 as part of a heterodimer with a direct GOLPH3 client. Alternatively, GALNT4 may be glycosylated itself and the global defects in glycosylation resulting from GOLPH3+3L deletion could result in its hypoglycosylation which may be detrimental to its stability. Attempts at a genetic rescue through overexpression of GOLPH3 only generated a partial recovery, most likely due to technical limitations of the experiment. Cells were harvested only 24 hours after transfection which is unlikely to allow sufficient time for the complete recovery of the clients. Furthermore, it is likely a proportion of the cells of the sample were not successfully transfected. These technical issues can be overcome by rescuing through stable genetic

integration of GOLPH3. A GOLPH3 mutant with alanine point mutations at conserved arginine residues was reported to ablate binding to coatomer (Tu et al., 2012). In conflict with this report, in the current thesis the mutant was able to pulldown coatomer from cell lysate and was able to rescue GALNT4 levels comparably to wild-type GOLPH3. Future biochemical and structural analysis will be required to clarify this discrepancy.

Phenotypes for the *ΔΔGOLPH3, GOLPH3L* cell line have been characterised from a triplet of robust flow-cytometry assays measuring different cellular features. It will be intriguing to use rescue experiments to test the functional dependency of GOLPH3 on PtdIns4P-binding, dimerisation, coatomer binding and cargo binding using GOLPH3 mutants for which these features are perturbed.

Deletion of GOLPH3 family genes caused a loss in tail-dependent Golgi retention, the lysosomal degradation of COPI cargo and global defects in glycosylation. In contrast, deletion of FAM114A genes caused no detectable phenotypes. Without a clear cellular phenotype is it difficult to establish a function. It is possible that the FAM114A family have a cell-type/condition-specific function which is redundant in U2OS cells under the current culture conditions. Of note, mutation of the trans-Golgi COPI scaffolding factor GORAB is associated with the progeroid disease gerodermia osteodysplastica (GO) (Hennies et al., 2008; Witkos et al., 2019). In a recent report, the molecular pathology was shown to involve defective COPI-dependent recycling of glycosylation enzymes resulting in concordant deficiencies in glycosylation (Witkos et al., 2019). Interestingly, phenotypes were less penetrant in HeLa cells than other cell types with a higher secretory demand such as fibroblasts (Witkos et al., 2019). Furthermore, despite wide-spread tissue expression of GORAB, GO manifests specifically in cell types which secrete a large amount of extracellular matrix proteins such as skin and osteoblasts (Hennies et al., 2008; Witkos et al., 2019). The glycosylation status of a cell is important for how it communicates with neighbouring cells and the extracellular environment. Relatively speaking, transformed cell lines in a monoculture are unlikely to be making a range of dynamic and varied interactions with their surrounding environment. This is in contrast to cells in the context of a tissue or an organism which interact with various different cell types and extracellular factors and whose interactions vary greatly during biological processes such as in development. How can different cell-types accommodate these varied and dynamic extracellular interactions? One strategy may be to modulate their glycosylation status. Perhaps under certain cellular conditions, FAM114A proteins are activated to alter the traffic of glycan-modifying enzymes in order to initiate a certain glycosylation programme and carry out a specific cellular function. The next chapter will begin to entertain these ideas by exploring the role of the FAM114A ortholog in the context of the model organism *Drosophila Melanogaster*.

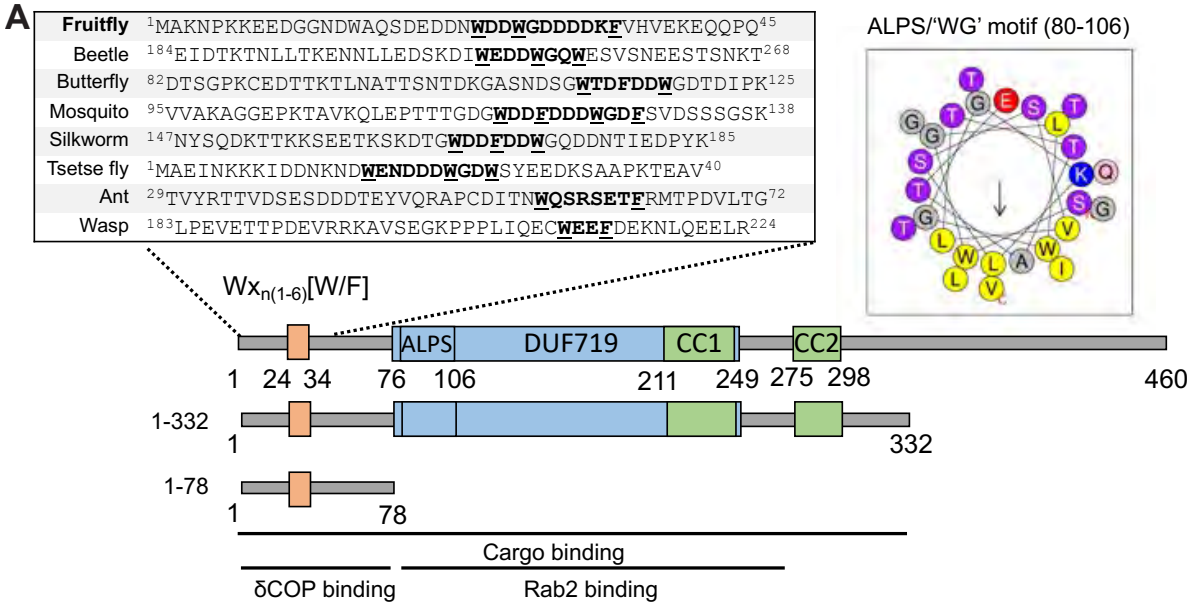
6 Orthologous *Drosophila* FAM114A protein CG9590 is a novel COPI adaptor for early intra-Golgi COPI vesicles

6.1 Introduction

FAM114A proteins specifically associate with early-Golgi membranes *in vivo* and FAM114A2 interacts with the cytoplasmic tails of COPI cargo *in vitro*. However, FAM114A knockout U2OS cell lines exhibit no clear or convincing phenotypes when tested in a number of different assays. As a consequence, very little understanding of the functional role of the FAM114A proteins could be attained. One possible explanation for the lack of a clear phenotype is that the FAM114A proteins have a cell-type-specific function which is not applicable in U2OS cells. It is possible to overcome this problem by deleting the necessary genes in the context of a whole organism with a diverse array of different cell types. As a result, we decided to switch from mammalian tissue culture cells to the model organism *Drosophila melanogaster* to investigate the function of FAM114A proteins.

6.2 The orthologous *Drosophila* FAM114A protein CG9590 interacts with the COPI coat and early-Golgi cargo

The fruit fly was deemed a favourable system to investigate FAM114A due to its tractable genetics and since only a single ortholog exists in flies which has already been partially characterised in S2 cells (Gillingham et al., 2014). The fly ortholog, *CG9590*, shares many similarities in sequence features to the mammalian orthologs (Figure 6.1). CG9590 has a DUF719 with a predicted ALPS motif at the N-terminus of the domain. It also has a region of predicted coiled-coil downstream of the DUF719 (Figure 6.1A). However, unlike the mammalian relatives, CG9590 has another region of predicted coiled-coil within the DUF719. Another key feature, which is conserved only amongst insects and not in mammals, is a predicted $Wx_{n(1-6)}$[W/F] motif which binds to the MHD of the δ-COP subunit of coatomer (Suckling et al., 2015). The $Wx_{n(1-6)}$[W/F] motif can accommodate 2 or 3 tryptophan or phenylalanine residues and is characteristically positioned within highly acidic stretches (Suckling et al., 2015). In summary, the human FAM114A proteins share high sequence conservation with their *Drosophila* relative (Figure 6.1B).



B

```
CG9590      1 MA----------------------KNPKKEEDGGNDW-AQSDEDDNWDDWGDDDDK----
HsFAM114A1  1 MSDDAGDTLATGDKAEVTEMPNSDSLPEDA-EVHCDSAAVSHEPTPADPRGEGHENAAVQ
HsFAM114A2  1 MS----------DKDDIETPLLTEAAPILE-DGNCEPAKNS--------------ESVDQ

CG9590     34 ---FVHVEKEQQPQEKVSSRNDAKGTAENQ---EKLE------TPRISAP--------AT
HsFAM114A1 60 GAGAAAIGPPVQPQDANALEPPLNGDVTEDTLAECIDSVSLEAEPRSEIPLQEQNYLAVD
HsFAM114A2 36 GAKPESKSEPVVSTRKR---P------------ETK--PSSDLETSKVLPIQDN----VS

CG9590     74 ASATSSSWGLWGGKLSSVLTTATEG---LGTITTQVNQGLNQIIGVPDPEELARINVAEK
HsFAM114A1 120 SPPSGGGWAGWGSWGKSLLSSASAT---VGHGLTAVKEKAGATLRIHGVNSGSSEGAQPN
HsFAM114A2  75 KDVPQTRWGYWGSWGKSILSSASATVATVGQQGISNVIEKAETSLGIPGPSEISTE-----

CG9590    131 --------AEKAAKAPPEENVDDEKESSAAFGL-----DFVTNLGSKVLNTGLDTLEGIG
HsFAM114A1 177 TENGVPEITDAATDQGPAES-PPTSPSSASRGMLSAITNVVQNTGKSVLTGGLDALEFIG
HsFAM114A2 130 -------VKYVAGETNAKEN-ENSSPVAGAFGVFSTISTAVQSTGKSVISGGLDALEFIG

CG9590    178 KKTMTILQDNDPKLMNKRKLLGLEANSPNLSAVLLEAKKEADQMEQSLQQLQLKRQKAQL
HsFAM114A1 236 KKTMNVLAESDPGFKRTKTLMER---TVSLSQMLREAKEKEKQRL--AQQLT-M--ERTA
HsFAM114A2 182 KKTMDVIAEGDPGFKRTKGLMNR---NATLSQVLREAKEKEEIRT--SNEVT-VETDKKT

CG9590    238 RFEILFENYCGLVHFEALEILSKESRLKLDTLLDAVSGNARAELQETINEVKELIELEDL
HsFAM114A1 288 HYGMLFDEYQGLSHLEALEILSNESESKVQSFLASLDGEKLELLKNDLISIKDIFAAKEL
HsFAM114A2 236 HYGLLFDEFQGLSHLEALEMLSQESEIKVKSILNSLSGEELETLKVELEQLKETFSLAEF

CG9590    298 DVESEGDYE-----AEELSERLAATIKDAELGIQFDDIANHWTQSLSWLTS---------
HsFAM114A1 348 ENEENQEEQGLEEKGEEFARMLTELLFELHVAATPDKLNKAMKRAHDWVEEDQTVVSVDV
HsFAM114A2 296 CEEEEEEKKG----DEDFTKDITELFSQLHVSSKPEKLARARNTAHEWIRKSLTKPLAEN

CG9590    344 -------------------------EEATAADLEQAYAKSIHALSEACALEMCKLHKIAE
HsFAM114A1 408 AKVSEEETKKEEKEEKSQDPQEDKKEEKKTKTIEEVYMSSIESLAEVTARCIEQLHKVAE
HsFAM114A2 352 -------------EEGEKQSEAENTEQVNKNSIEDIHAFAIRSLAELTACSIELFHKTAA

CG9590    379 LLLVKPHHSTAN--EVDGVVNLCQQFNGHLQGLSHRFAAVLSGKSE---TEESKGRVSTF
HsFAM114A1 468 LILHGQEEEKPAQDQAKVLIKLTTAMCNEVASLSKKFTNSLTTVGSNKKAEVLNPMISSV
HsFAM114A2 399 LVLHGRKQEVTAIERSQTLSQMTIVLCKELSSLSKEFTTCLTTAGVKEMADVLNPLITAV

CG9590    434 FSEMLSAVQFIEKAYKLFTPILQMGAV--------------------
HsFAM114A1 528 LLEGCNSTTYIQDAFQLLLPVLQVSHIQTSCLKAQP-----------
HsFAM114A2 459 FLEASNSASYIQDAFQLLLPVLEISLIENKIESHRHELQGQKPLLEH
```

Figure 6.1 The sequence features of orthologous *Drosophila* FAM114A protein CG9590. (**A**) A schematic displaying the domains of CG9590. A highly acidic $Wx_{n(1-6)}$[W/F] δ-COP MHD-binding motif (in bold, critical aromatic residues underlined) is conserved amongst the N-terminus of CG9590 insect relatives. Consistent with the mammalian orthologous proteins, HeliQuest analysis of the N-terminus of the DUF719 of CG9590 predicts the formation of an amphipathic lipid packing sensor (ALPS) motif (residues 80-106). Previous yeast-two hybrid studies have mapped residues 79-272 as a Rab2 binding site (Gillingham et al., 2014). (**B**) Clustal Omega multiple sequence alignment of orthologous *Drosophila* and human FAM114A proteins. Graphical alignments were generated using BoxShade.

As with the mammalian orthologous proteins, a proteomics approach was taken to characterise the binding partners for CG9590. GST-tagged CG9590 (1-332) was produced in bacteria and used to pull-out interactors from S2 cell lysate which were identified by mass spectrometry. CG9590 specifically bound to many of the components of the COPI coat, several COPI cargo proteins and components of the OST complex (Figure 6.2A, Supplementary Table 8.8). With the notable exceptions of the COPI coat and C16orf13, many of the interactions were conserved between CG9590 and human orthologous protein FAM114A2 (Chapter 4 and see section 6.4). Represented amongst the COPI cargo were a range of different families of proteins including glycan-modifying enzymes (pgants [GALNTs in mammals], FucTB, Ugt35a, oxt, Mgat1 and C1GalTA; Figure 6.3), p24 cargo receptors (Bai and p24-4), SNARE complex proteins (Sly1 and Sec22) and others (golgin-84, ZFPL1 and GLG1). As with the interacting partners of human FAM114A2, most of the interacting cargo proteins of CG9590 were predicted to reside in or around compartments at the ER/cis-Golgi interface. Since CG9590 was predicted to have a $Wx_{n(1-6)}$[W/F] motif and was able to pull-out coatomer from S2 cell lysate, this interaction was validated biochemically. To ablate the $Wx_{n(1-6)}$[W/F] motif, two key tryptophan residues were mutated to alanines (W24A, W27A). Similarly, two key residues were mutated in the *Drosophila* δ-COP MHD (H350A, K355S) which were predicted to ablate binding to $Wx_{n(1-6)}$[W/F] motif-containing proteins based on orthologous proteins in previous reports (Suckling et al., 2015). GST-tagged CG9590 truncations and 6xHis-tagged δ-COP MHD proteins were purified from bacteria for use in GST pulldowns to test direct binding. A truncated CG9590 (1-332) containing the $Wx_{n(1-6)}$[W/F] motif failed to bind to the δ-COP MHD while a shorter truncated form of CG9590 (1-78) exhibited strong binding (Figure 6.2B). It is plausible this discrepancy in binding is because the longer form of CG9590 is subject to autoinhibitory protein folding which is relieved by the truncation required to generate the shorter form of CG9590. This interaction was ablated when key residues were mutated in either the $Wx_{n(1-6)}$[W/F] motif of the short CG9590 truncation or in the δ-COP MHD. The wild-type δ-COP MHD exhibited no binding towards the GST negative control. Since CG9590 can bind to both COPI cargo and the COPI coat *in vitro*, it seems highly plausible that it functions as a COPI adaptor.

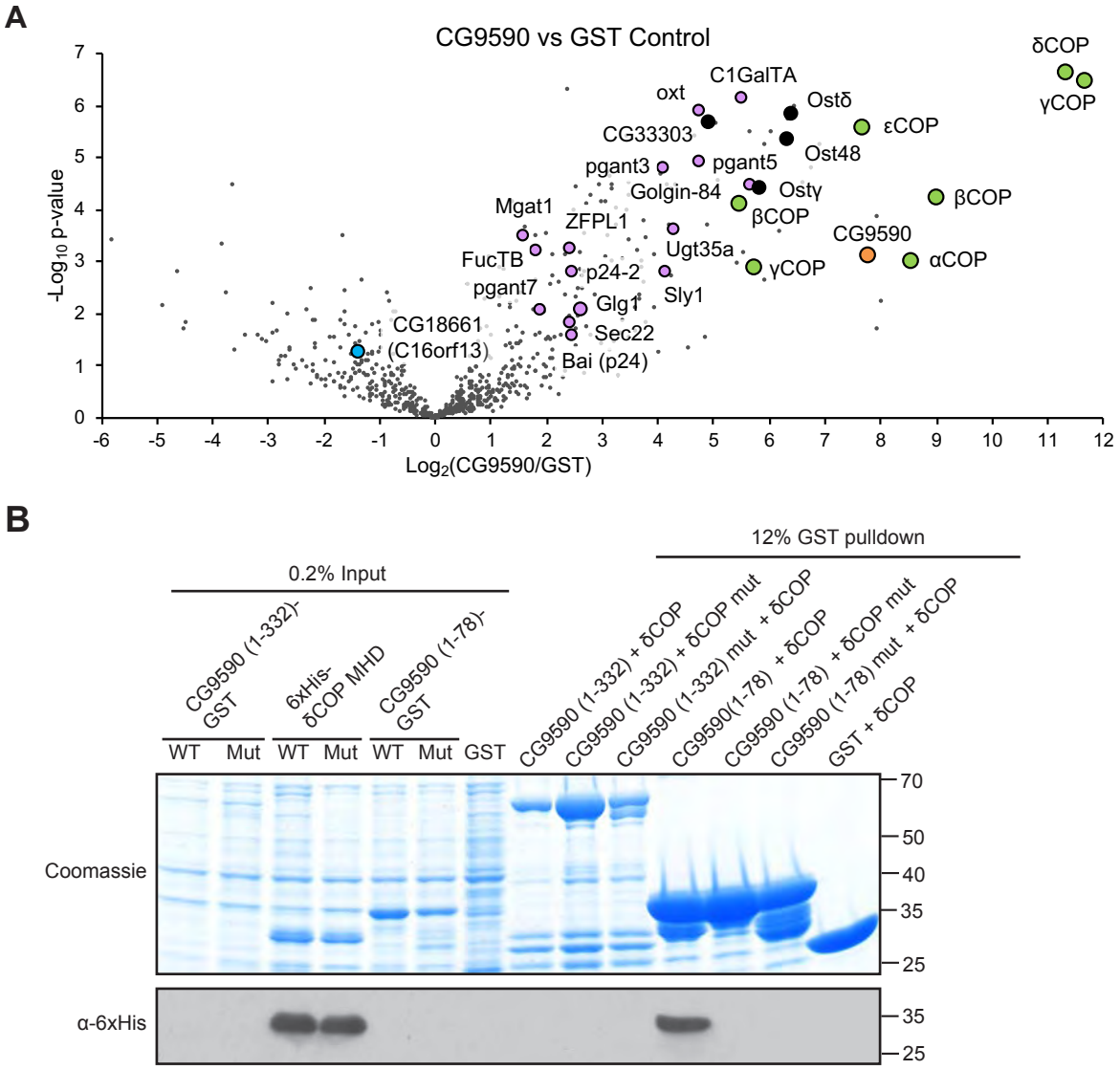


Figure 6.2 Orthologous FAM114A *Drosophila* protein CG9590 binds simultaneously to the COPI coat and COPI cargo *in vitro*. (**A**) Volcano plots comparing spectral intensity values generated from GST pulldowns from S2 cell lysate with GST-tagged CG9590 (residues 1-332) vs a GST control. -log P values were generated from Student's t-tests. Highlighted in green are subunits of the COPI coat, known COPI cargo or COPI-associated proteins in magenta, the *Drosophila* C16orf13 relative in blue, CG9590 in orange and components of the OST complex in black. Data was analysed using the Perseus platform and represent a triplicate of repeats. (**B**) Immunoblot and Coomassie stain from a GST pulldown testing direct binding between the δ-COP μ-homology domain (MHD, residues 291-532) and truncations of CG9590, both expressed in bacteria. Mutations in the $Wx_{n(1-6)}$[W/F] motif of CG9590 (WDDWGDDDDKF to ADDAGDDDDKF) and the δ-COP MHD (H350A, K355S) were included as they were predicted to ablate the interaction based off previous biochemical analyses using yeast and human proteins (Suckling et al., 2015). N=1.

As mentioned previously, many of the potential clients of CG9590 are cargoes which are expected to cycle within the early-Golgi or even recycle from the Golgi to the ER. With regards to the glycosylation enzyme clients of CG9590, this argument is supported by the fact they act early in the glycan-modification pathways to which they belong (Figure 6.3). A large proportion of the enzyme clients act in the initial steps of mucin-type O-linked glycosylation; the ppGalNAcTs (pgant3, pgant5 and pgant7 – GALNTs in mammals) catalyse the initial addition of N-acetylgalactosamine to a serine or threonine residue of a polypeptide (Figure 6.3A; Zhang and Ten Hagen, 2019). C1GalTs (C1GalTA) catalyse the secondary addition of N-acetylglucosamine to form the core 1 structure (Zhang and Ten Hagen, 2019). Thereafter, the core 1 structure undergoes extension during several further rounds of glycan modification. Another client associated with O-linked glycobiology, oxt, catalyses the addition of xylose to the serine residues of proteoglycans (Wilson, 2002). This precedes the considerable extension resulting from the conjugation of repeating glycosaminoglycan units. With regards to N-linked glycosylation, at the ER Alg3 is involved in one step of many in the synthesis of a dolichol-linked precursor which is transferred onto an asparagine of a nascent polypeptide by the OST complex (Figure 6.3B; Katoh and Tiemeyer, 2013). After glucosidase and mannosidase processing at the ER, the CG9590 client Mgat1 catalyses the first N-linked modification predicted to occur in the Golgi (Katoh and Tiemeyer, 2013). Biochemical and genetic data on FucTB is limited; but, based on close paralog FucTA (Fabini et al., 2001), it is likely to work later in the pathway to fucosylate N-linked glycans. The client Ugt35a is an orphan glycosyltransferase so predictions of its localisation in the early secretory pathway cannot be drawn from its position in a glycosylation pathway. In summary, many of the CG9590 enzyme clients operate in the very first or in very early glycan-modification steps with a particular bias toward mucin-type O-linked glycosylation. Future experiments are required to define the CG9590 consensus motif within the cytoplasmic tail of the clients.

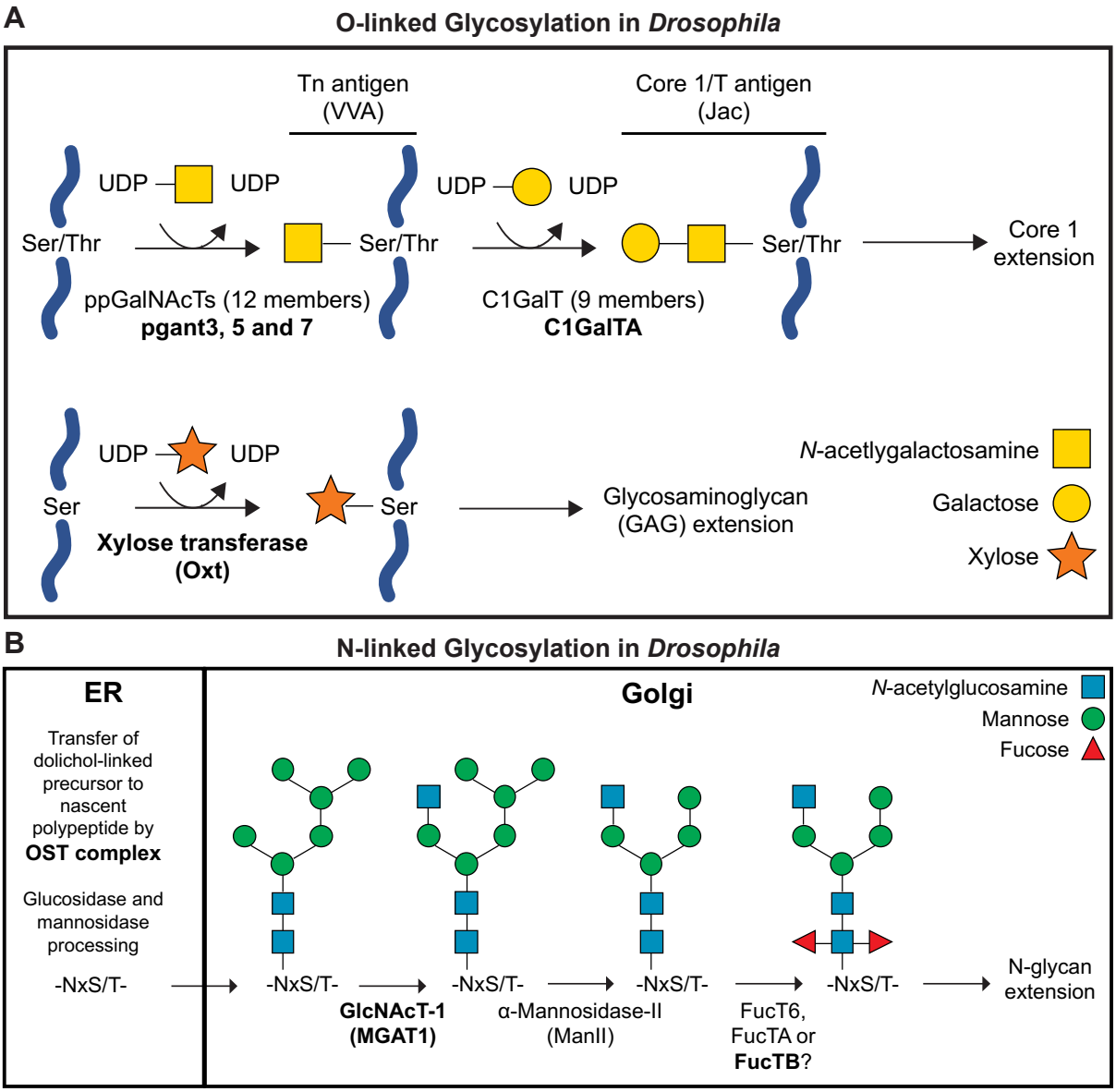


Figure 6.3 CG9590 enzyme clients act at early points in glycan-modifying synthetic pathways. A schematic displaying the position of CG9590 enzyme clients (in bold) in simplified O-linked (**A**) and N-linked (**B**) glycosylation pathways in *Drosophila*. Included are the antigen and lectin specificities (in brackets) of a pair of relevant glycans involved in mucin-type glycosylation.

6.3 *ΔCG9590* adult flies display a defect in O-linked mucin-type glycosylation

The evidence from the biochemical and proteomic analyses of CG9590 suggests that it is functioning as a COPI adaptor for vesicles cycling within the early-Golgi or between the Golgi and the ER. To explore this function in the context of an organism, CRISPR-Cas9 gene editing was utilised to delete the *CG9590* gene in the germline of the fly. The strategy was to target the 5' and 3' UTR of *CG9590* with the aim of deleting the whole open reading frame. Deletion of the whole gene allows confidence that a true null has been generated even in the absence of an antibody for validation purposes. Two independent *ΔCG9590* clones were generated which both displayed complete excision of the whole gene with subtle differences in the scarring at the point of re-ligation between the two cut sites (work done by Nadine Muschalik). Homozygous stocks of each line were viable with a subtle defect in viability relative to the parental wild-type line. To examine the glycosylation status of these lines, whole adult homozygous flies were homogenised and cell lysates were subject to lectin blotting using a panel of different fluorescein-conjugated lectins with a range of different sugar specificities (Figure 6.4). Of the lectins tested; DSL, ECL GSLII, LEL, STL and WFA probed blots revealed no striking differences between wild-type and knockout adult flies. DSL gave a very weak signal across all genotypes tested. In contrast to the other lectins, when blots were probed with fluorescein-conjugated VVA or Jacalin, a high molecular weight band (>260 KDa) appeared in both of the independent knockout samples but not in the wild-type sample (asterisks; Figure 6.4). A more striking difference was observed when blots were probed with fluorescein-conjugated Jacalin; the most abundant band in the wild-type lane was markedly depleted across both independent knockout lines (hash; Figure 6.4). Between them, these two lectins detect the linkages of the first two steps of mucin-type core 1 O-linked glycosylation (Figure 6.3). This is in striking accordance with the finding that several of the CG9590 enzyme clients are involved in these exact glycan modifications. This would suggest that deletion of CG9590 causes the mislocalisation of its enzyme clients thus triggering a downstream defect in glycosylation.

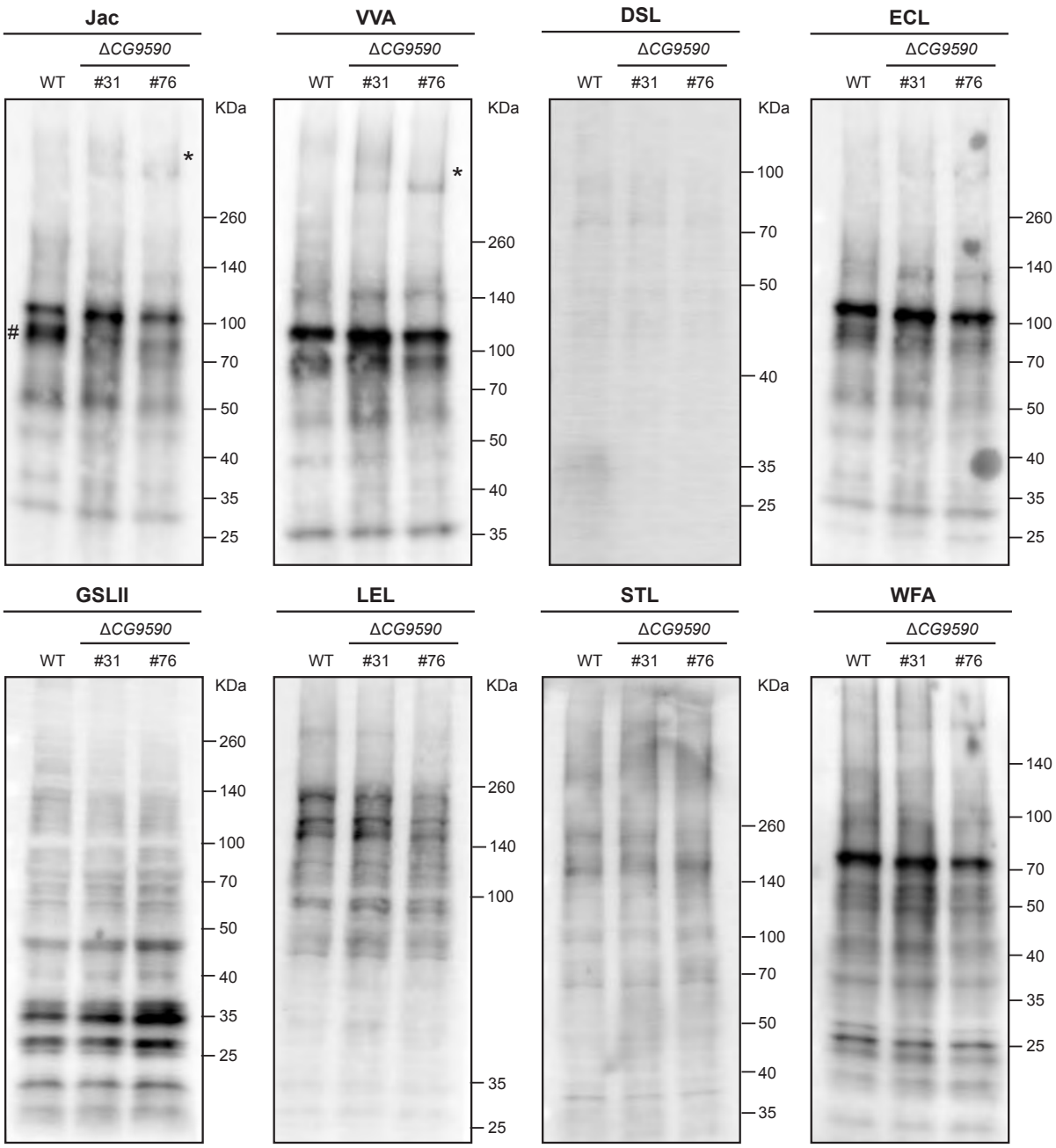


Figure 6.4 Lectin blot analysis reveals a defect in O-linked mucin-type glycosylation in homozygous *ΔCG9590* adult flies. Lectin blots of cell lysate from whole adult flies of a wild-type background and from two independent CRISPR-Cas9 knockout fly lines using a panel of fluorescein-conjugated lectins. The asterisks and hash in the Jac and VVA blots indicate protein bands which disappear or appear when comparing the *ΔCG9590* line and the wild-type control.

The species recognised by Jacalin which disappears after the deletion of *CG9590* represents one or more proteins which are approximately 70-100 KDa in size. It is likely these proteins are normally mucin-type O-glycosylated and are subsequently either hypoglycosylated and/or depleted when *CG9590* is deleted. To identify this species, lysates from adult flies were subject to tandem lectin affinity chromatography followed by protein identification by mass spectrometry. Proteins were first enriched based on their affinity for Jacalin which recognises the core 1 O-glycan disaccharide also known as T antigen (Figure 6.3A). Proteins from the flow-through were subsequently purified based on their affinity for VVA which recognises the glycan linkage known as the Tn antigen. The spectral count abundances of the proteins isolated from the two independent *ΔCG9590* lines were highly comparable so an average was taken. The average spectral counts of the proteins from the *ΔCG9590* sample were used to calculate a ratio in relation to the wild-type sample to generate a relative abundance score. Relevant proteins which were of moderate to high abundance in the wild-type sample, but showed considerable depletion in the knockout samples, are displayed in Table 6.3 (Supplementary Table 8.9). Across the whole data set, the protein enrichment for Jacalin and VVA was very similar when comparing within genotypes. Many of the proteins which exhibited a lower enrichment in the knockout samples were proteins of the secretory pathway and therefore likely to be glycosylated. The curated proteins displayed in Table 6.3 are all secretory or membrane proteins with the exception of Hts. Indeed, many of the proteins have been shown to be modified with mucin-type O-linked glycans (Mmp1, Gp93; Schwientek et al., 2007). Others are highly likely to be mucin-type glycosylated since they are mucins by definition (CG3777, Hmu) and/or contain regions which are rich in prolines, threonines and serines (PTS-rich regions) which are prone to mucin-type modification (CG10737, CG34396, CG44245). Common functional themes could be identified amongst the different proteins which were depleted in the knockout samples. Many proteins were identified as secreted peptidases or glycanases (especially chitinases) which are important in extracellular remodelling processes. Another group of proteins were categorised as proteins secreted by fat body cells such as fat body protein 1 and components of the larval serum protein 1 complex. A number of proteins had undefined functions and/or no clear human ortholog upon which to make functional predictions. In summary, deletion of *CG9590* causes the hypoglycosylation and/or depletion of a range of different mucin-type O-glycosylated proteins in the adult fly. This is likely to be an indirect consequence of the mislocalisation of the glycosyltransferase clients as a result of the deletion of *CG9590*. The list of hypoglycosylated and/or depleted proteins will help inform forthcoming detailed phenotypic analyses of the *ΔCG9590* fly lines.

Table 6.3

Gene Name	KDa	Relative Abundance Δ*CG9590*/WT		Description
		Jacalin	VVA	
Cda5	205	0.13	0.29	Chitin deacetylase
Cht5	67	0.15	0.28	Secreted chitinase
Cht7	113	0.18	0.16	Secreted chitinase
Mmp1	61	0.36	0.22	Metallopeptidase
Hexo1	71	0.00	0.00	Plasma membrane glucosidase
CG3777	94	0.16	0.48	Sequence homology with muc5AC and mucin-19 in related Drosophila species
Hmu	64	0.38	0.23	Hemomucin
Lsp1alpha	99	0.00	0.02	Larval serum protein 1 complex, secreted fat body protein.
Lsp1beta	96	0.00	0.00	
Lsp1gamma	93	0.09	0.21	
Fbp1	120	0.05	0.13	Fat body protein 1, secreted
CG10737	102	0.27	0.14	PTS rich, integral membrane protein
CG34396	111	0.39	0.25	PTS rich, integral membrane protein
CG44245	50	0.00	0.03	PTS rich
CG7896	134	0.13	0.46	Integral membrane protein
Gp93	90	0.17	0.29	ER and secreted O-linked glycoprotein protein
Hts	82	0.12	0.16	Adducin orthologous protein, peripheral membrane protein

A curated list from mass spectrometry analysis of tandem lectin affinity chromatography of cell lysate from whole adult flies. Lysates were subject to Jacalin affinity chromatography and the flow through was subject to VVA affinity chromatography. Relative abundance was generated as a ratio of the total spectral counts from the wild-type sample and the average of the spectral counts from two independent *ΔCG9590* samples. Shown are functionally relevant proteins which show the biggest decrease in enrichment in the knockout samples vs the wild-type sample. Proteins are categorised according to function: peptidases and glycanases (green), mucins (blue), fat body proteins (orange), miscellaneous (grey).

6.4 Discussion

CG9590 simultaneously pulled-down the COPI coat and early-Golgi/ER COPI cargo suggesting it is likely to be a novel COPI adaptor for early intra-Golgi COPI vesicles. Many of the interactions were conserved between *Drosophila* CG9590 and human FAM114A2 including interactions with the ppGalNAcTs (GALNTs and pgants), the OST complex, p24 cargo receptors, ATP synthase, fatty acid transporters and elongation factor 2 (Table 6.4.1). The fact that these *in vitro* interactions were conserved between the orthologous proteins provides added confidence that they are valid. However, proteins such ATP synthase reside in the inner membrane of the mitochondria and are therefore likely to represent a spurious interaction. Conserved interactions of FAM114A proteins with fatty acid transporters and translational proteins were all neglected in this thesis but will be the subject of future studies. The power of *Drosophila* genetics can be leveraged to ascertain if there are any phenotypes which are relevant to the interacting proteins in question.

Table 6.4.1

FAM114A2 Interactor/Human Relatives	**CG9590 Interactor/*Drosophila* Relatives**
GALNT2, GALNT3, GALNT12	pgant3, 5, 7
RPN2, DDOST	OstΔ, Ost48, CG33303
TMED2, TMED10	p24-2, bai
CYC1	Cyt-c1
ATP5D, ATP5C1	ATPsyndelta, ATPsyn-gamma, ATPsynF/CG4692, ATPsyn-b, ATPsynO/oscp, ATPsynbeta
SLC27A1, SLC27A2	Fatp1
eIF-2gamma	eIF-2gamma, eIF-2alpha

Shown are the conserved interactors between orthologous human and *Drosophila* FAM114A proteins. FAM114A2 interactors correspond to Figure 4.1 and Supplementary Table 8.2, CG9590 interactors correspond to Figure 6.2 and Supplementary Table 8.8.

A notable interaction that was not conserved in human cells was the interaction of CG9590 with the COPI coat. CG9590 was shown to interact with the MHD of the δ-COP subunit of the COPI coat through a $Wx_{n(1-6)}$[W/F] motif at its N-terminus (Suckling et al., 2015). The $Wx_{n(1-6)}$[W/F] motif in CG9590 is only conserved amongst insects and may explain why mammalian FAM114A proteins do not pull-down coatomer. Vps74 was shown to interact with multiple subunits of coatomer simultaneously (Tu et al., 2008). It remains to be seen if CG9590 interacts with other COPI subunits in addition to δ-COP. It is conceivable that another CG9590-coatomer interface exists which is conserved in mammals. The reason to why such an interaction is elusive in the mammalian context is not clear. Perhaps the

FAM114A proteins in mammals have a conditional function and under steady-state conditions exist in an inhibited conformation which blocks binding to the COPI coat. Interestingly, of the CG9590 truncations containing the $Wx_{n(1-6)}$[W/F] motif, the short fragment (1-78) bound directly to the δCOP MHD unlike the larger fragment (1-332). Despite this, the larger fragment was still capable of pulling out the whole COPI coat from S2 cell lysate. This could be indicative of a number of things. It is possible the larger fragment was incapable of binding to the δCOP MHD due to it existing in an autoinhibitory folded state. It is plausible this autoinhibitory fold is removed by truncation in the case of the shorter fragment. Despite this, why was the larger fragment capable of pulling out the whole COPI coat? One reason could be that CG9590 is interacting with coatomer through another interface involving other coat subunits as mentioned previously. Another idea is that the larger fragment can bind to coatomer in the presence of cargo-detergent micelle complexes. Perhaps only when CG9590 docks onto the cargo tails and the micelle (mimicking a lipid-bilayer), is autoinhibition relieved allowing CG9590 to interact with the MHD of δCOP.

Further validation of the function of CG9590 as a COPI adaptor was that homozygous *ΔCG9590* adult flies displayed a global defect in O-linked mucin-type glycosylation which was concordant with the putative CG9590 clients. Deletion of GOLPH3 family members causes their clients to mislocalise to the lysosome for degradation (Schmitz et al., 2008; Tu et al., 2008; Chapter 5). It is probable that the deletion of *CG9590* also causes the mislocalisation and degradation of the glycan-modifying enzyme clients causing downstream defects in glycosylation. Lectin blot analysis of lysate from whole adult flies using a panel of lectins only revealed subtle defects in mucin-type glycosylation but not in any of the other glycan types tested. This was despite CG9590 interacting with enzymes operating in a variety of different glycosylation pathways. Since lectin blots were carried out using lysate from whole adult flies, it is likely some tissue- and temporal-specific glycosylation defects were masked. Different glycan-modifying enzymes have been shown to be important for different cells-types and at different developmental stages (Hagen et al., 2009). Of the O-linked mucin-type glycosylation enzymes in *Drosophila*, there are 12 members of the pgant family and 9 members of the C1GalT family with the different individual enzyme knockouts displaying tissue-specific phenotypes (Zhang and Ten Hagen, 2019). Since CG9590 specifically interacted with C1GalTA and pgant3, 5 and 7; it will be interesting to see if the *CG9590* knockout phenocopies the individual enzyme knockout flies. It will be valuable to conduct different lectin stains on different tissues to understand in what cell-types *CG9590* is important. This cell-type specific function may be conserved in mammals and could therefore be leveraged in the hunt for the elusive mammalian phenotype.

As mentioned previously, lectin blot analysis of whole adult flies revealed only subtle mucin-type differences between wild-type flies and the two *CG9590* lines. Surprisingly, lectin affinity chromatography resulted in the identification of numerous relevant proteins of different sizes whose enrichment was vastly decreased in both knockout fly lines. Several of these proteins have been previously shown to be mucin-type O-glycosylated in S2 cells which validates the relevance of the lectin affinity data set (Schwientek et al., 2007). The discrepancy between the lectin blot and lectin affinity chromatography suggests that lectin blotting of whole adult flies is not the most sensitive method for detecting glycosylation defects. Lectins recognising other sugar-types will be revisited using different detection modalities in future studies. Many of the proteins which were reduced in lectin affinity knockout samples were chitinases and fat body proteins. Chitinases are important for remodelling the extracellular matrix therefore it will be interesting to see if the *CG9590* knockout flies exhibits corresponding defects (Pesch et al., 2016).

It is important to consider that many mucins are heavily glycosylated (carbohydrates can account for 90% of the weight of the macromolecule) and secreted mucins assemble into very large oligomeric structures resulting in the formation of insoluble gels *in vivo* (Ferez-Vilar and Hill, 1999). In the lectin affinity enrichment process, adult flies were solubilised in a non-reducing aqueous buffer in the presence of detergent and the lysate was clarified by centrifugation. It is highly likely that many of these gel-forming mucins would have remained insoluble under these conditions and therefore would have been lost from analysis. Enriched proteins were subsequently resolved by SDS-PAGE using a 4-20% Tris-Glycine gel. For the mucins that were successfully enriched by lectin affinity, it was likely they were poorly resolved by SDS-PAGE due to their extreme size and physiochemical nature. Bespoke agarose gel electrophoresis could be used for future analyses (Ramsey et al., 2016; Rodríguez-Piñeiro et al., 2012). And finally, for the proportion of mucins that were soluble and well resolved, due their highly-glycosylated nature they may be resistant to trypsin digestion and will be troublesome to identify by mass spectrometry (Godl et al., 2002). In summary, it is likely that current methods for measuring mucin glycosylation are suboptimal and are therefore greatly underrepresenting the discrepancies in glycosylation between the wild-type and *ΔCG9590* lines.

Human FAM114A2 appeared to recognise the cytoplasmic tails of enzymes with polybasic membrane-proximal stretches (Chapter 4). Since many of the interactions with enzymes are conserved between humans and *Drosophila*, it is conceivable that CG9590 interacts with cytoplasmic tails through a similar motif. No recognition motif is immediately clear when comparing the cytoplasmic tails of CG9590 clients; but some clients may be binding indirectly as part of a heterodimer or through the COPI coat (Table 6.4.2).

Table 6.4.2

Gene	Cytoplasmic Tail	Topology
pgant3	MGLRFQQLKKLW	Type II
pgant5	MTFSTFTRKMRGRMRSNTCR	Type II
pgant7	MRVSTIRSGRICR	Type II
C1GalTA	MTANSLLGRSILNEGRSNKRS	Type II
oxt	MEQSVSARWLKRYR	Type II
Mgat1	MRTRK	Type II
FucTB	MRLAQR	Type II
Ugt35A	QRVFVSLISKIRKNRAEASKLKTQ	Type I
p24-2	NLFHTCVYINEQCLCVCCLYIKKISALGLINDKINGLPDP	Type I
Bai (p24)	YLRRYFKAKKLIE	Type I
Glg1	RATKRAMGLKNK	Type I
Ost48	KDEPVGRAAKEDKKSQ	Type I
CG33303	DFSITSHAHKE	Type I

Shown are the cytoplasmic tails of putative CG9590 clients. Clients with large cytoplasmic regions are not shown. Cytoplasmic tail sequences are assigned according to Uniprot. Proteins of type II topology are aligned right and those of a type I topology are aligned left.

The *in vivo* Golgi retention assay, the lysosomal degradation assay and lectin stains were robust assays used to successfully characterise the phenotypes associated with the deletion of GOLPH3 family members in U2OS cells (Chapter 5). Moving forward, the same assays could be adapted for use in conjunction with *CG9590* RNAi in S2 cells. It will be interesting to track the localisation of clients and the glycosylation status of specific tissues at specific stages after *CG9590* deletion. Notably, Yamamoto-Hino and colleagues have generated a transgenic line expressing HA-tagged Mgat1 which would be intriguing to cross into the *ΔCG9590* background to see if the enzyme is mislocalised and/or degraded (Yamamoto-Hino et al., 2012).

In summary, CG9590 can interact with the COPI coat and early-Golgi COPI cargo *in vitro*. Preliminary analysis reveals that deletion of *CG9590* causes a defect in mucin-type O-linked glycosylation in adult flies which is concordant with a proposed mislocalisation and/or degradation of CG9590 glycan-modifying enzyme clients. Based on these findings, I propose that CG9590 is a novel COPI cargo adaptor which defines a subclass of vesicles which recycle from early-Golgi cisternae. Based on the high conservation of primary sequence and *in vitro* binding partners, this function of CG9590 is likely to be conserved across orthologous eukaryotic proteins. By selectively targeting a subset of early-acting glycan-modifying enzymes, CG9590 functions to fine tune the localisation of Golgi enzymes across the Golgi stack to contribute the fidelity of glycosylation within the cell.

7. Conclusions and perspectives

7.1 Updated models for Golgi enzyme trafficking

At the start of this thesis, a series of complimentary models were suggested describing the trafficking mechanisms of Golgi enzymes. With the initial aim of better understanding the diversity and nuances of intra-Golgi transport, this thesis has unearthed some novel insights into these trafficking processes. In this chapter, I will attempt to re-evaluate and complement the existing models in light of the new progress.

7.2 FAM114A – a novel family of early-Golgi COPI adaptors which fine tune the sub-Golgi localisation of glycan-modifying enzymes

In an attempt to better understand intra-Golgi trafficking processes, a MitoID screen was employed to identify novel intra-Golgi vesicle-resident proteins. The screen identified FAM114A2, an evolutionary conserved Rab2 effector of unknown function (Gillingham et al., 2014; Gillingham et al., 2019). FAM114A2, paralogous protein FAM114A1 and *Drosophila* orthologous protein CG9590 were all shown to specifically associate with early-Golgi membranes (Gillingham et al., 2014). FAM114A proteins share a predicted ALPS motif which is necessary for membrane association, presumably through absorbing onto highly-curved membranes. *In vitro*, FAM114A2 and CG9590 were shown to specifically pulldown a number of COPI cargo proteins associated with the early-Golgi and ER. Furthermore, CG9590 was demonstrated to interact directly with the COPI coat through a well-defined and validated motif. Finally, the deletion of *CG9590* caused a mucin-type O-linked glycosylation defect in adult flies which corresponded with a predicted mislocalisation and/or depletion of CG9590's putative clients. With those findings in mind, I propose that FAM114A proteins represent a novel family of evolutionary conserved early-Golgi COPI adaptors. I suggest that FAM114A proteins are specifically recruited to early-Golgi cisternae where they select a specific subset of cargo destined for retrograde COPI-dependent recycling to the cis-Golgi, ERGIC and possibly even the ER (see Figure 7.1A). I propose that since GOLPH3+3L are highly promiscuous COPI adaptors which are specifically recruited to the trans-Golgi by PtdIns4P, they function as master gatekeepers at the late-Golgi to retain a vast array of Golgi residents. In contrast, by virtue of its early-Golgi-targeting and selectivity, FAM114A proteins specifically recycle a smaller subset of cargo earlier in the Golgi to generate an asymmetry in the distribution of Golgi residents across the stack. This provides a means for the fine-tuning of the sub-Golgi localisation of glycan-modifying enzymes. The compartmentalisation of glycan-modifying enzymes, broadly in accordance with their position in their enzymatic pathway, serves to ensure the fidelity of cellular glycosylation.

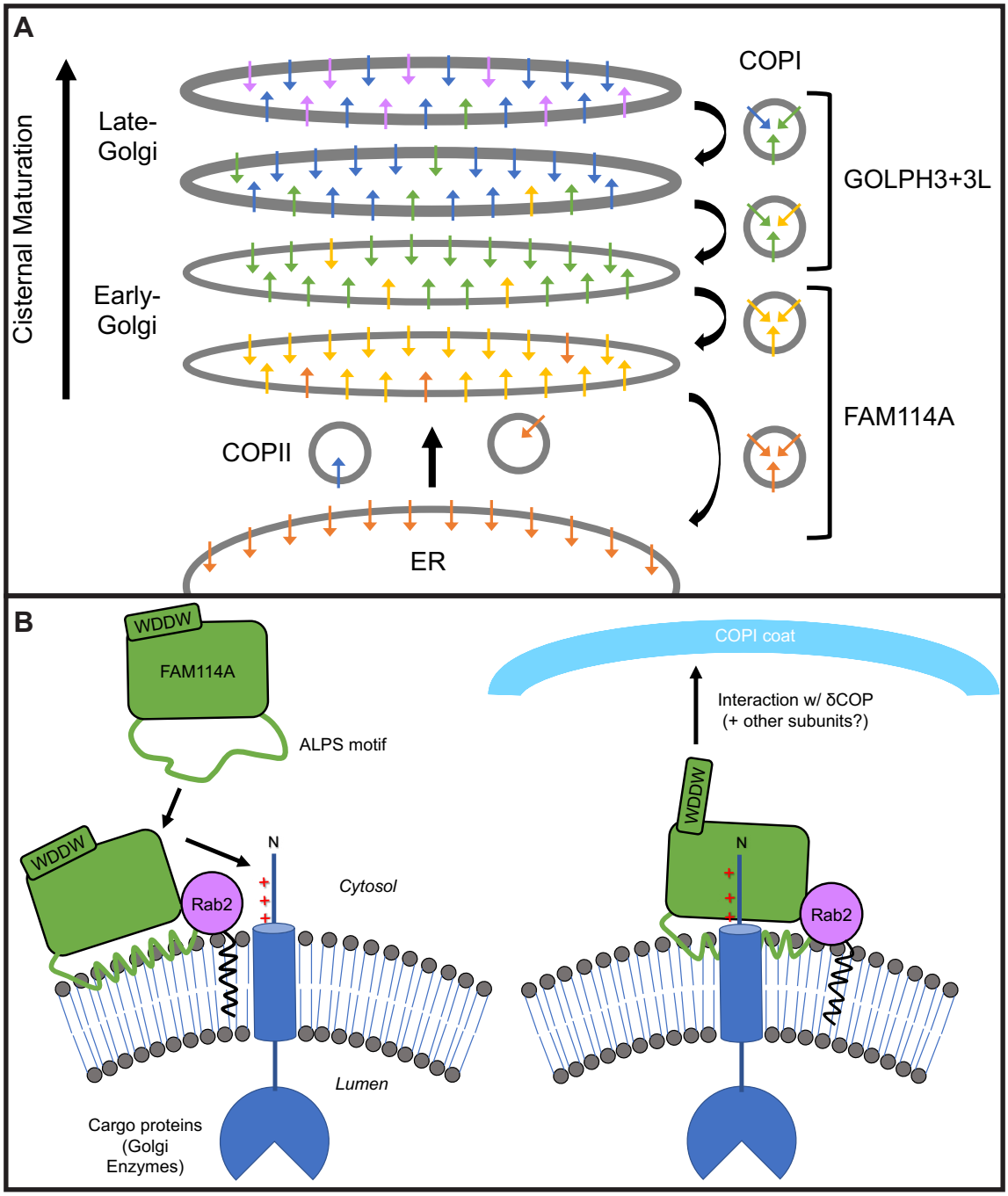


Figure 7.1 Speculative model of FAM114A function. (**A**) I propose that FAM114A family proteins act as COPI vesicle adaptors which operate on a specific subset of vesicles and cargo at the early-Golgi. In selecting a specific subset of cargo, FAM114A proteins serve to fine tune the distribution of Golgi-resident cargo throughout the stack. In contrast, GOLPH3 proteins function as promiscuous COPI adaptors at the late-Golgi. They act as master gatekeepers at the late-Golgi to retain a vast array of different of Golgi-residents. (**B**) FAM114A proteins are likely to absorb onto highly-curved membranes of budding vesicles through a predicted ALPS motif. It is likely Rab2 is required to specifically recruit FAM114A to the early-Golgi. Here, FAM114A can sample the membrane-proximal positive-charges in the cytoplasmic tails of its client. In insects, once FAM114A is docked onto membranes and its clients, it can associate with the MHD of the δ-COP subunit of the COPI coat through a $Wx_{n(1\text{-}6)}[W/F]$ motif. It remains to be seen if FAM114A can interact with other subunits of the COPI coat in insects, or at all in other eukaryotic life.

What are the molecular mechanisms which govern this proposed function of FAM114A family proteins? Without structural insights it is difficult to draw strong conclusions; however, it is possible to speculate based on the biochemical and cell biology data. It is conceivable that in the cytosol, FAM114A proteins exist in an autoinhibited state in which they cannot interact with cytosolic coatomer or positive charges on disordered stretches of proteins which may resemble recognition motifs in the cytoplasmic tails of clients (Figure 7.1B). FAM114A proteins absorb onto highly-curved Golgi membranes through their ALPS motif, perhaps after sufficient curvature has been initiated in a nascent COPI bud by Arf1-GTP (Beck et al., 2008). It is likely that FAM114A proteins are specifically recruited to highly-curved membranes of the early-Golgi by Rab2-GTP; however, it is plausible that other factors could facilitate this specific membrane association. Once docked onto membranes, the FAM114A proteins can sample the cytoplasmic tails of its clients and simultaneously interact with the COPI coat to promote the segregation of cargo into the budding vesicle.

Within insects, orthologous CG9590 proteins can interact with the MHD of the δ-COP subunit of coatomer through a $Wx_{n(1-6)}$[W/F] motif at their N-terminus. It remains to be seen if CG9590 can interact with other subunits of coatomer and if any of these coatomer interactions are conserved in humans. Attempts to demonstrate an *in vitro* interaction between human FAM114A proteins and coatomer have been unsuccessful. Furthermore, deletion of FAM114A genes in transformed human cell lines have failed to result in any convincing detectable phenotypes. In contrast, *ΔCG9590* homozygous adult flies display a defect in mucin-type O-glycosylation. Notably, the ability of CG9590 to bind to mucin-type glycan-modifying enzymes is shared with human FAM114A2. On that basis, it is plausible that human FAM114A proteins may be of functional relevance in a specialised mucin-secreting cell-type. Mucin-secreting cells represent a cell-type with an extreme glycosylation requirement. They often produce vast amounts of mucins which are extremely heavily glycosylated and are loaded into space-consuming secretory granules which crowd the cytoplasm of the cell (Birchenough et al., 2015; Ferez-Vilar and Hill, 1999). Under such extreme demands, it is conceivable that the cell may have bespoke mechanisms to increase its mucin-type glycosylation capacity and it would therefore be interesting to explore FAM114A proteins in this context. Of note, a novel mechanism was proposed by the Bard laboratory which claims that under activation by growth factors, GALNTs are relocated from the Golgi to the ER causing an upregulation in mucin-type O-glycosylation (Gill et al., 2010). While the reproducibility and validity of the GalNAc-T activation (GALA) pathway has been questioned (Chia et al., 2019; Herbomel et al., 2017), preliminary

findings suggest that GALNT2 is localised at the ER in the mucin-secreting HT-29 cell line and this will form the basis of future investigation (data not shown).

Interestingly, while fungi have a *GOLPH3* ortholog - *Vps74* in budding yeast, they lack a FAM114A ortholog. This could perhaps be explained by the relative simplicity of the glycobiology of fungi compared to that of metazoans. In contrast, plants lack a *GOLPH3* ortholog but have a FAM114A ortholog. Work is currently ongoing to characterise the *Arabidopsis* ortholog of FAM114A and it will be interesting to see if FAM114A in this context has evolved to compensate for the lack of a *GOLPH3* ortholog through a relative expansion of its client list.

7.3 The adaptor competition model for the differential distribution of Golgi-residents

GOLPH3+3L were shown to pulldown a large proportion of the Golgi-resident glycan-modifying enzymes of the cell, suggesting they function as promiscuous adaptors for COPI vesicles budding from the trans-Golgi. Putative GOLPH3+3L clients include the early-acting enzyme GALNT2 and the late-acting enzyme ST6GAL1 (Eckert et al., 2014; Isaji et al., 2014). The *in vivo* Golgi retention assay revealed that a chimeric reporter containing the cytoplasmic tail of GALNT2 was more efficiently retained in the Golgi by GOLPH3 proteins than the ST6GAL1 cytoplasmic tail reporter. A proposed model is that during COPI vesicle budding at the trans-Golgi, multiple clients could compete for binding to GOLPH3+3L and those with the strongest binding affinities will be more readily packaged into vesicles for retrograde recycling (Figure 7.2A). In theory, this would generate an asymmetry in the distribution of Golgi-residents between the neighbouring cisternae. Based on simple biochemical comparisons of the binding affinities of the GALNT2 and ST6GAL1 chimeras, no clear difference between binding affinities were observed either way (Chapter 4, Figure 4.6). Quantitative biophysical analysis of GOLPH3-cytoplasmic tail binding affinities and higher-resolution characterisation of the relative sub-Golgi localisation of the chimeras are required to properly challenge this model.

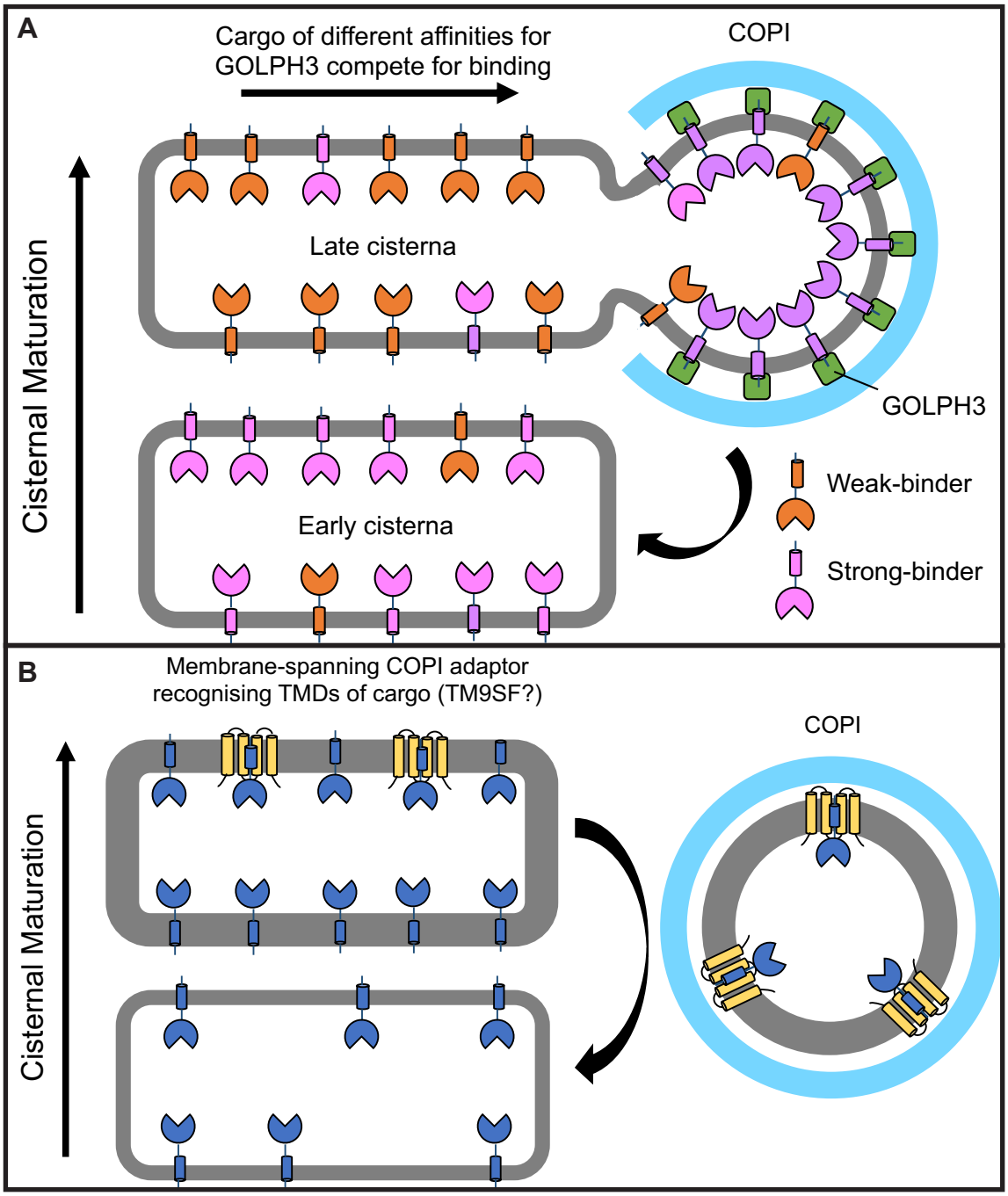


Figure 7.2 Further models of Golgi enzyme retention. The leading preliminary models for the retention of Golgi enzymes were given at the start of this thesis in Figure 1.3. Represented are some additional complimentary models based on the findings of the current thesis and other recent reports. (**A**) The adaptor competition model for the differential distribution of Golgi residents. GOLPH3+3L were shown to be promiscuous COPI adaptors operating at the trans-Golgi. It is possible that different clients with different affinities for GOLPH3 compete with each other for binding. The clients with the strongest binding affinities are more readily recycled in GOLPH3-dependent COPI vesicles leading to an asymmetry in the distribution of Golgi-residents between cisternae. (**B**) Membrane-spanning COPI adaptor model for Golgi retention. It is plausible that there are membrane-spanning COPI adaptors that simultaneously interact with the COPI coat and specifically recognise the TMDs of Golgi enzymes to help package them into COPI vesicles. A combination of recent reports suggest that the TM9SF family, particularly TM9SF2, can interact with coatomer and the TMDs of Golgi enzymes.

7.4 The membrane-spanning COPI adaptor model for Golgi retention

The *in vivo* Golgi retention assay was not just leveraged to explore cytoplasmic tail-dependent mechanisms of Golgi retention, but also to interrogate TMD-dependent mechanisms. According to the membrane thickness model, Golgi enzymes are retained at the Golgi based on the physiochemical properties of their relatively short TMDs (Bretscher and Munro, 1993; Sharpe et al., 2010). Only 3 TMDs were tested in the *in vivo* Golgi retention assay (GALNT2, ST6GAL1, sucrase-isomaltase) so strong conclusions on the membrane thickness model cannot be drawn either way; however, the data are not fully concordant with the model. As expected based on previous reports, a sucrase-isomaltase plasma membrane reporter with its relatively-long TMD (20 amino acids) exhibited no Golgi retention while a chimeric reporter containing the short TMD of ST6GAL1 (17 amino acids) exhibited robust Golgi retention (Liu et al., 2018; Munro, 1991). The apparent departure from the membrane thickness model came when a chimeric reporter containing the short TMD of GALNT2 (17 amino acids) did not exhibit Golgi retention. Despite two reporters containing TMDs of equal predicted length, they behaved differently *in vivo*. It is plausible that there are some subtle features of the physiochemical properties of the TMDs which can explain this disparity. An alternative and perhaps complimentary model is that membrane-spanning COPI adaptor(s) exist which simultaneously interact with the TMDs of the enzymes and coatomer to package the enzymes into budding COPI vesicles (Figure 7.2B). TM9SF proteins are an evolutionary-conserved family of multi-pass integral membrane proteins which have been shown to interact directly with coatomer and the TMD of Golgi enzymes and their deletion causes global defects in glycosylation (Pacheco et al., 2018; Tanaka et al., 2017; Tian et al., 2018; Woo et al., 2015; Yamaji et al., 2019). Experiments are ongoing to test the capacity of the cell to retain the ST6GAL1 TMD chimera after the deletion of TM9SF genes. ST6GAL1 was shown to exhibit at least two mechanisms of Golgi-retention: a TMD-dependent mechanism and a cytoplasmic tail-, GOLPH3-dependent mechanism. Why would an enzyme require two mechanisms for retention? It is possible that this simply represents built-in redundancy or a feature to improve the efficiency with which Golgi enzymes are packaged into vesicles and recycled. Alternatively, by having two parallel trafficking mechanisms it could provide a means to fine tune the sub-Golgi localisation of different enzymes.

A large scale bioinformatic analysis of bitopic membrane proteins revealed that Golgi-residents have a shorter TMD than their post-Golgi counterparts (Sharpe et al., 2010). The analysis did not identify any clear potential sequence specific retention motifs in the TMDs of Golgi-residents; but, it is possible such a signal would be lost due to the vast nature of

the data set (Sharpe et al., 2010). Recent work in plants have identified a specific glutamine reside in the TMD of a Golgi enzyme which is important for retention (Schoberer et al., 2019). It will be intriguing to identify the factor(s) responsible for this retention and to see if such a mechanism is conserved across eukaryota. Since a common feature of many Golgi-residents is their short TMD, could it be that the length of the TMD is an important feature for recognition by a membrane-spanning Golgi retention factor? In other words, could a long TMD evade such a retention factor and therefore serve as an exit signal?

7.5 Models for the lysosomal-targeting of mislocalised Golgi enzymes

What other roles could the short TMDs of Golgi-residents have? Since their short size is an evolutionary-conserved feature, it is highly likely to be of cellular significance. In yeast, plants, mammals and insects; the perturbation of Golgi-retention mechanisms triggers the vacuolar/lysosomal-degradation of the mislocalised Golgi enzymes (Chang et al., 2013; Schmitz et al., 2008; Schoberer et al., 2019; Tu et al., 2008). What remains a mystery is the mechanism by which these enzymes are targeted to the endolysosomal system to be turned-over. It is conceivable that with the perturbation of retrograde COPI-dependent recycling mechanisms, Golgi enzymes will begin to accumulate in the thicker sphingolipid/sterol-rich bilayer of the late-Golgi (Figure 7.3). The prolonged hydrophobic mismatch of the short TMDs in the incompatibly thick late-Golgi bilayer could somehow serve as a cue for lysosomal targeting and degradation. One possibility is that prolonged mismatch could cause the TMDs to begin to aggregate together which could serve as a signal for degradation. But what carriers are responsible for transporting these mislocalised enzymes to the lysosome? The GGAs are a family of clathrin adaptors which are involved in the generation of carriers budding from the TGN destined for the endolysosomal system (Uemura and Waguri, 2020). Good evidence to support the idea that lysosomal-targeting of mislocalised Golgi enzymes is GGA-dependent is that in yeast, *Vps74* has a negative genetic interaction with the GGAs (Wood et al., 2012). It is plausible that when both GGAs and *Vps74* are deleted, the cell can no longer turn-over mislocalised enzymes leading to the accumulation of aggregates at the late-Golgi and/or aberrant glycosylation due to the incorrect compartmentalisation of enzymes.

But how do GGAs recognise their cargo? Some operate through receptors while others have been shown to interact with ubiquitinated cargo (Uemura and Waguri, 2020). One possibility is that a membrane-spanning adapter recognises the mismatched-TMD of a Golgi enzyme and simultaneously interacts with the GGAs (Figure 7.3). The receptor may be itself ubiquitinated as a means to recruit the GGAs or as a downstream trafficking signal rather than a degradation signal. Alternatively, the mismatched-TMD of a Golgi enzyme could trigger the recruitment of an E3 ligase to ubiquitinate lysine residues or the terminal amide group of the cytoplasmic tails of the mislocalised enzymes. This could be a membrane-spanning Golgi-localised E3 ligase similar to the yeast protein Tul1 or a cytosolic ligase recruited indirectly by another means (Reggiori and Pelham, 2002). The resulting poly-ubiquitin chain could then serve to recruit the GGAs which would subsequently direct the enzymes for lysosomal degradation. With the establishment of a quantitative flow-cytometry based lysosomal degradation assay in *ΔΔGOLPH3, GOLPH3L*

U2OS cells, it is now possible to test if a mismatched/mislocalised-short TMD is a cue for their degradation. Furthermore, this assay in combination with genetic tools could be leveraged to screen for the factors involved in such a pathway. Preliminary spatial proteomics analysis comparing wild-type and *ΔΔGOLPH3, GOLPH3L* U2OS cells confirms that a number of Golgi-resident proteins are shifted to the lysosome in knockout cells (data not shown). Interestingly, the analysis also revealed that a pair of E3 ligases are specifically recruited to the Golgi from the cytosol when *GOLPH3* and *GOLPH3L* are deleted. Furthermore, the multipass integral membrane protein LAPTM4B shifted from the endosome to the Golgi, a protein which is known to shuttle between the Golgi and the lysosome in a GGA- and ubiquitin-dependent manner (Meng et al., 2016). Could these factors be key components of this trafficking pathway?

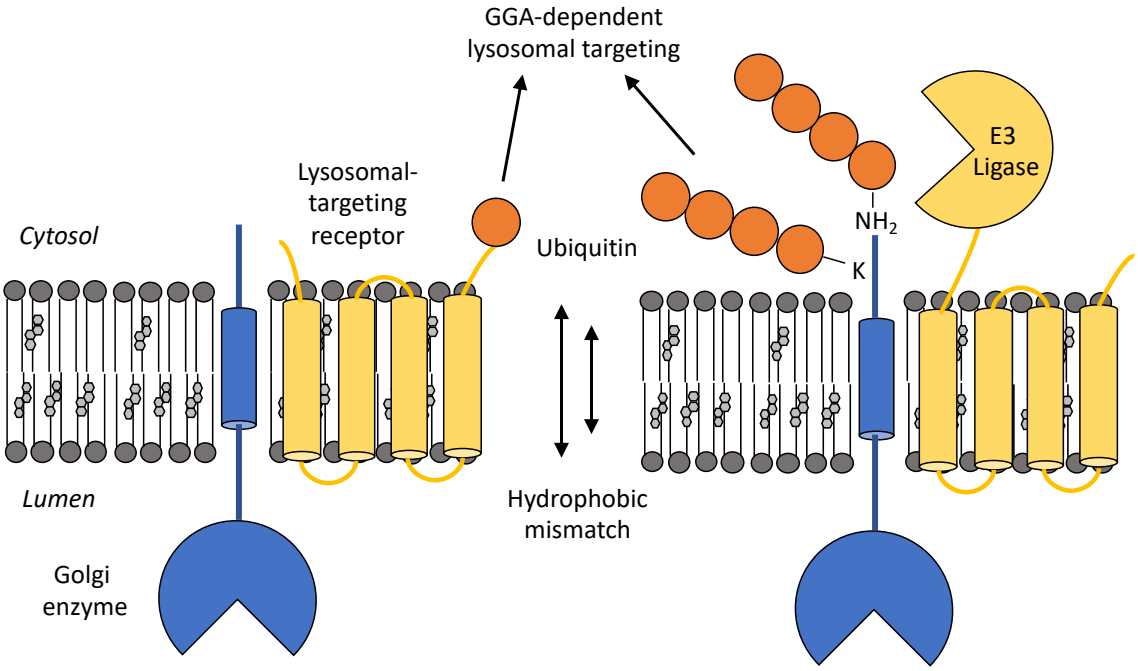


Figure 7.3 Models for the lysosomal-targeting of mislocalised Golgi enzymes. In the event of the pertubation of Golgi-retention mechanisms, it is possible that the Golgi enzymes begin to accumulate in late-Golgi membranes. It is possible the prolonged hydrophobic mismatch of the short TMDs of Golgi enzymes in the thick sphingolipid- and sterol-rich membranes of the late-Golgi serve as a cue for lysosomal-targeting and degradation. The mismatched-TMDs could be recognised by a membrane-spanning receptor which may target the enzyme for lysosomal degradation via a GGA-dependent pathway. It is conceivable that such a receptor may be ubiquitinated to allow recognition by the GGAs. Alternatively, the mismatched-TMDs could trigger the recruitment of a ubiquitin E3 ligase (membrane-spanning or cytosolic) to polyubiquitinate the lysine residues and/or terminal amine groups of the cytoplasmic tails of the enzymes. The polyubiquitin signal would serve as a cue for the recruitment of GGAs and subsequent lysosomal-targeting and degradation.

7.6 Conclusions

At the beginning of this thesis, I set out to explore the nuances of intra-Golgi traffic with the aim of understanding how Golgi-resident glycan-modifying enzymes are differentially distributed across the Golgi stack. In light of my own findings and the recent work of others, it has become increasingly clear that the mechanisms of intra-Golgi vesicular cargo selection are more diverse than previously described. For many years, investigating mechanisms of intra-Golgi traffic had been limited by the technical difficulties of resolving neighbouring cisternae by light microscopy. Perhaps with the progression of super-resolution light microscopy, expansion microscopy and high-resolution spatial proteomics, it will become easier to characterise the sub-Golgi distribution of different enzymes (Kurokawa et al., 2013; Parsons et al., 2019; Wassie et al., 2019). Furthermore, armed with a combination of CRISPR-Cas9 genetic tools and the trilogy of *in vivo* flow-cytometry-based assays described in this thesis, it will be possible to rigorously test and identify novel Golgi enzyme trafficking mechanisms in the future.

8. Appendix

Presented in the appendix are the tabulated data sets associated with the proteomic analyses discussed in this thesis. With the exception of Supplementary Table 8.7 and 8.9, all tables represent the results from Student's t-tests comparing the spectral intensity of proteins between two conditions from GST affinity chromatography experiments. Tables are truncated to display the proteins with the most statistically significant differences. Supplementary Table 8.7 shows the tabulated form of a Venn diagram comparing Supplementary Tables 8.1, 8.2 and 8.5. Below is legend for the supplementary tables.

Supplementary Table (Page)	Title	Associated Figure/Table (Page)
8.1 (127)	FAM114A1-GST vs GST	Figure 4.1A (67)
8.2 (128)	FAM114A2-GST vs GST	Figure 4.1B (67)
8.3 (129)	FAM114A1 (wild-type)-GST vs FAM114A1 'WG' mutant-GST	Figure 4.1A (67)
8.4 (129)	FAM114A2 (wild-type)-GST vs FAM114A2 'WG' mutant-GST	Figure 4.1B (67)
8.5 (130)	GST-GOLPH3, GST-GOLPH3L vs GST	Figure 4.2A (68)
8.6 (137)	GST-GOLPH3 vs GST-GOLPH3L	Figure 4.2B (68)
8.7 (140)	COPI proteome, GOLPH3+3L and FAM114A2 interactome Venn diagram	Figure 4.2C (68)
8.8 (149)	GST-CG9590 (1-332), CG9590-GST (1-332) vs GST	Figure 6.2A (105)
8.9 (151)	Lectin affinity chromatography *ΔCG9590* vs wild-type adult flies	Table 6.5 (111)

Supplementary Table 8.1. FAM114A1-GST vs GST

Protein Name	Gene Name	-Log_{10} p-value	Difference Log_2(FAM114A1/GST)
Protein NOXP20	FAM114A1	3.546908	12.485611
Ras-related protein Rab-2A	RAB2A	3.59420963	9.32901891
Ras-related protein Rab-2B	RAB2B	3.43145612	5.85614777
Elongation factor 1-beta	EEF1B2	1.31302091	5.48168182
Elongation factor 1-gamma	EEF1G	1.07879732	4.34225909
Small integral membrane protein 7	SMIM7	0.98997924	4.26590919
Beta-actin-like protein 2	ACTBL2	0.89242803	3.6971674
Ig alpha-1 chain C region	IGHA1	1.09297061	3.67886925
Cysteine--tRNA ligase, cytoplasmic	CARS	1.61514972	3.51000341
Stress-70 protein, mitochondrial	HSPA9	1.24053968	3.06954638
UPF0585 protein C16orf13	C16orf13	3.12134601	3.06394895
ADP/ATP translocase 2	SLC25A5	0.87865442	3.02892939
Heparan sulfate 2-O-sulfotransferase 1	HS2ST1	1.04826476	2.74117215
LanC-like protein 2	LANCL2	0.94875385	2.71778488
Thioredoxin	TXN	0.87505809	2.69340452
Elongation factor 1-alpha 1	EEF1A1	0.96293311	2.02686564
Carbonyl reductase [NADPH] 3	CBR3	1.13760926	1.75858688
Peptidyl-prolyl cis-trans isomerase A	PPIA	1.33579712	1.75509834
Profilin-1	PFN1	1.21530727	1.5873057
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	1.14444513	1.2600282
39S ribosomal protein L40, mitochondrial	MRPL40	0.92661147	1.2211628
ATP synthase subunit gamma, mitochondrial	ATP5C1	2.32024611	1.1974837
Carbonyl reductase [NADPH] 1	CBR1	1.43731508	0.87119039

Supplementary Table 8.2. FAM114A2-GST vs GST

Protein name	Gene name	-Log10 p-value	Difference Log2(FAM114A2/GST)
Protein FAM114A2	FAM114A2	1.66633971	9.78488795
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	0.98951417	6.55590566
OCIA domain-containing protein 1	OCIAD1	0.87452268	6.16647339
Long-chain fatty acid transport protein 4	SLC27A4	1.15069141	5.69500287
Fatty aldehyde dehydrogenase	ALDH3A2	0.97854872	5.6706454
Heparan sulfate 2-O-sulfotransferase 1	HS2ST1	0.88199925	5.10686239
Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	B3GAT3	0.83640578	5.03785133
Elongation factor 1-beta	EEF1B2	1.19226224	4.97987239
Mannosyl-oligosaccharide glucosidase	MOGS	1.03631069	4.86892064
Nuclease EXOG, mitochondrial	EXOG	1.00702322	4.76409658
Type 1 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	TMEM55B	1.26076254	4.73175049
Chondroitin sulfate glucuronyltransferase	CHPF2	0.9252304	4.71513494
Golgi membrane protein 1	GOLM1	1.11113669	4.54028702
Cytoskeleton-associated protein 4	CKAP4	0.81223871	4.46262678
Glutaminyl-peptide cyclotransferase-like protein	QPCTL	0.78858148	4.39885775
Protein O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase 1	POMGNT1	1.10712701	4.38753064
ATP synthase subunit delta, mitochondrial	ATP5D	1.57909893	4.37726212
Selenoprotein H	SELH	0.90899216	4.27469063
ATP synthase subunit gamma, mitochondrial	ATP5C1	1.33457589	4.14327558
Protein YIPF6	YIPF6	1.29649409	4.08930969
Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	MGAT2	0.81012052	4.08677165
UDP-glucuronic acid decarboxylase 1	UXS1	1.05500414	4.07487361
Polypeptide N-acetylgalactosaminyltransferase 3	GALNT3	0.77895379	4.05774244
ATP-dependent zinc metalloprotease YME1L1	YME1L1	0.77893894	4.05308787
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	2.43321447	3.93559837
Beta-galactosidase-1-like protein 2	GLB1L2	0.99261665	3.80536397
Cysteine--tRNA ligase, cytoplasmic	CARS	1.68107519	3.685112
LEM domain-containing protein 2	LEMD2	0.83837784	3.59936142
Catechol O-methyltransferase domain-containing protein 1	COMTD1	0.76406643	3.58228556
Polypeptide N-acetylgalactosaminyltransferase 12	GALNT12	1.35428353	3.54470984
ADP/ATP translocase 2;ADP/ATP translocase 2, N-terminally processed	SLC25A5	1.00959724	3.5402298
Glycosaminoglycan xylosylkinase	FAM20B	0.7786641	3.48929024
Cytochrome c1, heme protein, mitochondrial	CYC1	0.73619466	3.48561541
78 kDa glucose-regulated protein	HSPA5	0.9392707	3.46965027
Elongation factor 1-gamma	EEF1G	0.84500418	3.43676313
Delta(24)-sterol reductase	DHCR24	0.93409413	3.33874512
Microsomal glutathione S-transferase 3	MGST3	1.77580922	3.21915181
LanC-like protein 2	LANCL2	1.1219401	3.19605637
MICOS complex subunit MIC60	IMMT	0.79452556	3.11462593
Heat shock cognate 71 kDa protein	HSPA8	0.8008276	3.0038929
Transmembrane protein 143	TMEM143	0.80261746	2.99981944
UPF0585 protein C16orf13	C16orf13	2.91187015	2.89078712
Stress-70 protein, mitochondrial	HSPA9	1.06332063	2.864151
Glycoprotein endo-alpha-1,2-mannosidase-like protein	MANEAL	0.9204624	2.76336098
Chromodomain-helicase-DNA-binding protein 4	CHD4	0.7353302	2.69585546
Long-chain fatty acid transport protein 1	SLC27A1	0.80411846	2.69186401
Nuclear mitotic apparatus protein 1	NUMA1	0.99640567	2.63863118
Transmembrane emp24 domain-containing protein 2	TMED2	1.26539313	2.55371157
ATP-binding cassette sub-family D member 3	ABCD3	0.71740202	2.4244798
Cytochrome c oxidase assembly factor 3 homolog, mitochondrial	COA3	0.71985791	2.39910634
Transmembrane emp24 domain-containing protein 10	TMED10	0.70818252	2.36066182
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	0.7913114	2.33496793
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	0.78713974	2.27886772
40S ribosomal protein S20	RPS20	0.89049269	2.26393509
60S ribosomal protein L18a	RPL18A	0.83377964	2.25886091
Lamin-B1	LMNB1	1.0310275	2.18209775
Heat shock-related 70 kDa protein 2	HSPA2	1.02942961	2.11260478
Protein-tyrosine sulfotransferase 2	TPST2	1.02616314	2.05196762
Carbonyl reductase [NADPH] 3	CBR3	1.1778569	1.99933497

Type 2 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	TMEM55A	0.76358868	1.91581726
Plasminogen receptor (KT)	PLGRKT	1.01547152	1.88610967
Protein transport protein Sec61 subunit beta	SEC61B	0.7816466	1.78751628
Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	0.70898091	1.77955945
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	1.06085752	1.75939624
Elongation factor 1-alpha 1	EEF1A1	0.72518187	1.67488861
SRSF protein kinase 1	SRPK1	0.98352604	1.66927528
Peptidyl-prolyl cis-trans isomerase A;	PPIA	0.77939859	1.47222837
39S ribosomal protein L48, mitochondrial	MRPL48	1.20667886	0.78861046
HIG1 domain family member 1A, mitochondrial	HIGD1A	1.37146519	0.51385434

Supplementary Table 8.3 FAM114A1 (wild-type)-GST vs FAM114A1 'WG' mutant-GST

Protein name	Gene name	-Log_{10} p-value	Difference Log_2(FAM114A1 WT/'WG' mutant)
Ig alpha-1 chain C region	IGHA1	1.28462295	4.07175255
Dermcidin;Survival-promoting peptide;DCD-1	DCD	1.21858083	3.58691216
Beta-actin-like protein 2	ACTBL2	1.25286099	3.46683947
Interferon-induced guanylate-binding protein 1	GBP1	1.2268496	3.34582074
Elongation factor 2	EEF2	1.10363633	2.2130661
Vacuolar protein sorting-associated protein VTA1 homolog	VTA1	1.98704941	2.168046
Superkiller viralicidic activity 2-like 2	SKIV2L2	1.24071723	1.93321101
U2 small nuclear ribonucleoprotein B	SNRPB2	1.04320881	1.6216526
28S ribosomal protein S5, mitochondrial	MRPS5	1.54827472	1.55227343
60S ribosomal protein L6	RPL6	0.96542891	1.4546833
Polyadenylate-binding protein 2	PABPN1	1.22395615	1.37655576
D-glucuronyl C5-epimerase	GLCE	1.33606215	1.17858728
Single-stranded DNA-binding protein, mitochondrial	SSBP1	1.0189153	1.11311913
Nucleoside diphosphate kinase, mitochondrial	NME4	2.24096929	1.072052
ATP synthase subunit a	MT-ATP6	1.04068441	1.02072779
Elongation factor 1-gamma	EEF1G	1.26800063	0.97572263
X-ray repair cross-complementing protein 5	XRCC5	1.06186208	0.45754178
40S ribosomal protein S26	RPS26	1.37848479	0.4493084
Polypeptide N-acetylgalactosaminyltransferase 12	GALNT12	1.0675704	-0.3560429
Interleukin enhancer-binding factor 3	ILF3	1.05612396	-0.5257772
60S ribosomal protein L13	RPL13	1.48211023	-0.970698
DNA replication licensing factor MCM3	MCM3	1.02278738	-1.0146268
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	1.3684095	-1.264445
GrpE protein homolog 1, mitochondrial	GRPEL1	1.365662	-1.4441261
Splicing factor 3B subunit 6	SF3B6	1.19506071	-1.8329713
Glycosaminoglycan xylosylkinase	FAM20B	1.07946555	-1.974665
60S ribosomal protein L9	RPL9	1.12908551	-2.0485306
Histone H1x	H1FX	1.3238244	-2.4526361
Probable ATP-dependent RNA helicase DDX17	DDX17	1.3066603	-2.9601949

Supplementary Table 8.4 FAM114A2 (wild-type)-GST vs FAM114A2 'WG' mutant-GST

Protein name	Gene name	-Log_{10} p-value	Difference Log_2(FAM114A2 WT/'WG' mutant)
Nucleoprotein TPR	TPR	0.84554737	3.2939593
60S ribosomal protein L10a	RPL10A	1.99731259	3.15402603
Zinc finger RNA-binding protein	ZFR	0.95376643	3.11507702
Methionine--tRNA ligase, cytoplasmic	MARS	1.24008203	2.90257517
Histone H2A.V/Z	H2AFV/Z	0.95694005	2.58337402
ATP-dependent RNA helicase DDX50	DDX50	1.14181999	2.56764158
Probable ATP-dependent RNA helicase DDX27	DDX27	1.0930571	2.5255963
60S ribosomal protein L7a	RPL7A	0.89633193	2.35910924
Polyadenylate-binding protein 4	PABPC4	1.06582971	2.14748446
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	1.05707648	1.872612
Exosome complex component RRP45	EXOSC9	1.07196405	1.79365921
40S ribosomal protein S13	RPS13	0.97273443	1.70287959
Probable ATP-dependent RNA helicase DDX5	DDX5	0.96729752	1.66949844
ATP-dependent RNA helicase DDX39A/B	DDX39A/B	0.83458116	1.66778501
Lamin-B1	LMNB1	1.92154916	1.63444265
Zinc finger protein 512	ZNF512	0.90575957	1.62879149
DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H	1.25126816	1.55527306
60S ribosomal protein L37a	RPL37A	1.1292183	1.5047245

Dynein light chain 1/2, cytoplasmic	DYNLL1/2	1.60009752	1.48327764
Polyadenylate-binding protein 1	PABPC1	1.19098248	1.46122106
39S ribosomal protein L41, mitochondrial	MRPL41	1.3359878	1.37078762
60S ribosomal protein L12	RPL12	1.01787688	0.75089963
40S ribosomal protein S8	RPS8	0.96542146	0.65847333
39S ribosomal protein L48, mitochondrial	MRPL48	1.35011532	0.55388896
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	0.94565482	-0.4532655
Kinesin-like protein KIF14	KIF14	0.83273938	-1.4953162
Dystrophin	DMD	0.83917579	-2.293197
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	0.91287821	-3.0131855
Small nuclear ribonucleoprotein Sm D3	SNRPD3	0.9207965	-3.3318507

Supplementary Table 8.5 GST-GOLPH3, GST-GOLPH3L vs GST

Protein name	Gene name	-Log10 p-value	Difference Log2 (GOLPH3+3L/ GST)
Aspartate--tRNA ligase, mitochondrial	DARS2	6.02003491	15.4368823
Deleted in autism protein 1	C3orf58	7.37893274	14.0743774
Zinc finger protein-like 1	ZFPL1	6.46208496	13.7011989
Cytochrome c oxidase assembly factor 3 homolog, mitochondrial	COA3	6.34938701	13.4899076
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	6.20094687	13.4819523
Mannosyl-oligosaccharide glucosidase	MOGS	4.04795006	13.4588121
Leucine-rich repeat-containing protein 59	LRRC59	7.30903732	13.382274
Mitochondrial import inner membrane translocase subunit TIM14	DNAJC19	5.12414344	13.3177118
Mitochondrial import inner membrane translocase subunit TIM16	PAM16	5.89789971	13.2240601
SUN domain-containing protein 2	SUN2	7.03754945	13.2185917
Monoacylglycerol lipase ABHD12	ABHD12	5.92787381	12.6696456
Transmembrane protein 206	TMEM206	5.99750814	12.6112677
Serine palmitoyltransferase 2	SPTLC2	5.50779898	12.4298916
Nucleolar complex protein 4 homolog	NOC4L	6.43796811	12.2818727
GPI transamidase component PIG-S	PIGS	5.57275277	12.2665113
Peroxisomal membrane protein PEX16	PEX16	6.68196444	12.225606
2-hydroxyacylsphingosine 1-beta-galactosyltransferase	UGT8	4.67116168	12.2042128
GPI-anchor transamidase	PIGK	6.92973857	12.1463591
Syntaxin-12	STX12	5.6384398	12.1299795
Telomere-associated protein RIF1	RIF1	5.10936214	12.1172469
Transmembrane protein 201	TMEM201	5.26630847	12.1152178
Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	MAN1B1	6.3637801	12.0430555
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	NDUFB10	7.51857748	12.0264498
Coatomer subunit beta	COPB1	5.79744472	11.9506118
Protein FAM3C	FAM3C	5.76994684	11.9198329
Testis-expressed sequence 10 protein	TEX10	6.30791314	11.8616323
Transmembrane protein 192	TMEM192	6.00227653	11.8453592
Glutaminyl-peptide cyclotransferase-like protein	QPCTL	6.23787941	11.8364979
FAS-associated factor 2	FAF2	5.37985756	11.7372096
LETM1 domain-containing protein 1	LETMD1	5.02600135	11.6446749
Extended synaptotagmin-2	ESYT2	4.28313002	11.6261066
BRI3-binding protein	BRI3BP	5.16179594	11.6032575
Serine palmitoyltransferase 1	SPTLC1	5.70024302	11.5890169
Phosphatidate cytidylyltransferase, mitochondrial	TAMM41	5.55486715	11.558273
Dynamin-like 120 kDa protein, mitochondrial	OPA1	7.35569511	11.516573
AFG3-like protein 2	AFG3L2	4.76554454	11.4516093
Nucleolar protein 14	NOP14	6.16373915	11.4379813
Prohibitin	PHB	5.96406637	11.2872661
Coatomer subunit gamma-2	COPG2	4.36366539	11.2871081
Ganglioside-induced differentiation-associated protein 1	GDAP1	4.83479157	11.2756116
Mitochondrial import receptor subunit TOM70	TOMM70A	4.79909197	11.2071247
Nucleolar complex protein 2 homolog	NOC2L	4.0684994	11.1496261
Sigma non-opioid intracellular receptor 1	SIGMAR1	7.40849318	11.066954
LETM1 and EF-hand domain-containing protein 1, mitochondrial	LETM1	4.49501205	11.0628993
N-acetylgalactosaminyltransferase 7	GALNT7	4.48802105	11.0302229
2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3	OGFOD3	7.02499314	11.0274951
Sulfhydryl oxidase 2	QSOX2	5.94520859	10.9209633
Beta-1,4-glucuronyltransferase 1	B4GAT1	6.31476756	10.9132163
Golgi SNAP receptor complex member 1	GOSR1	4.66653785	10.8934317
Vesicle-trafficking protein SEC22b	SEC22B	3.87197803	10.8544251
Calcineurin B homologous protein 1	CHP1	6.65834643	10.8453906
Dehydrogenase/reductase SDR family member 13	DHRS13	5.41347744	10.8427022
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6	NDUFB6	4.85487557	10.8385057
Protein CASP	CUX1	5.51380633	10.8198713
Transmembrane protein 109	TMEM109	4.79818746	10.7605728
Polypeptide N-acetylgalactosaminyltransferase 6	GALNT6	4.96599029	10.6913878
Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	MGAT2	4.64152794	10.6461159
Transmembrane protein 97	TMEM97	6.84625735	10.6460959
Myb-binding protein 1A	MYBBP1A	3.87104014	10.6301314
Protein transport protein Sec61 subunit beta	SEC61B	5.23241539	10.5997477
Ubiquitin-protein ligase E3C	UBE3C	3.00639262	10.5160535
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial	NDUFB5	5.19491774	10.4393355
Vesicle transport through interaction with t-SNAREs homolog 1B	VTI1B	4.67706082	10.3985078
Sphingosine-1-phosphate lyase 1	SGPL1	3.63813829	10.3823716
ADP/ATP translocase 2;ADP/ATP translocase 2, N-terminally processed	SLC25A5	2.95277233	10.2986139
CCAAT/enhancer-binding protein zeta	CEBPZ	5.35324212	10.2833392
ATP-binding cassette sub-family B member 10, mitochondrial	ABCB10	3.5644912	10.2810726
Sphingomyelin phosphodiesterase 4	SMPD4	4.14594312	10.2416484
Cytochrome c1, heme protein, mitochondrial	CYC1	7.12317668	10.2399022
Nuclear pore complex protein Nup205	NUP205	3.91040365	10.2102076
Dolichol-phosphate mannosyltransferase subunit 1	DPM1	4.24238774	10.200497

Uncharacterized protein NCBP2-AS2	NCBP2-AS2	5.17117187	10.1555901
Glucoside xylosyltransferase 1	GXYLT1	5.22402213	10.1551781
Golgi apparatus protein 1	GLG1	3.06875944	10.1502021
Translational activator GCN1	GCN1L1	3.03545396	10.1050116
Mitochondrial inner membrane protein OXA1L	OXA1L	5.03160522	10.1036323
ATPase family AAA domain-containing protein 3A	ATAD3A	4.62563277	10.0651531
ADP-ribosylation factor-like protein 6-interacting protein 1	ARL6IP1	6.37787098	10.040501
Catechol O-methyltransferase domain-containing protein 1	COMTD1	6.21273939	9.98005168
Pachytene checkpoint protein 2 homolog	TRIP13	4.94766063	9.96813774
Calcium-independent phospholipase A2-gamma	PNPLA8	5.16983194	9.96137746
Alpha-(1,6)-fucosyltransferase	FUT8	4.52053285	9.92963219
Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	3.9060679	9.9254837
Cytochrome c oxidase subunit NDUFA4	NDUFA4	4.7445388	9.9075098
Coatomer subunit epsilon	COPE	6.82980987	9.90577443
Polypeptide N-acetylgalactosaminyltransferase 3	GALNT3	4.50877475	9.87219588
Ephrin-B1	EFNB1	5.035337	9.86801656
Armadillo repeat-containing X-linked protein 3	ARMCX3	5.17190897	9.8431282
SET domain-containing protein 9	SETD9	4.94677429	9.77752622
Heparan sulfate 2-O-sulfotransferase 1	HS2ST1	5.52203144	9.77539031
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	3.50314611	9.77337392
Mitochondrial import receptor subunit TOM20 homolog	TOMM20	4.61926322	9.70804755
Optic atrophy 3 protein	OPA3	4.52265843	9.63193544
4F2 cell-surface antigen heavy chain	SLC3A2	5.75561059	9.62290923
CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1	ST3GAL1	4.99338475	9.61640644
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	MAN1A2	3.69898187	9.53474808
Protein transport protein Sec61 subunit alpha isoform 1	SEC61A1	5.58239133	9.5209837
Fukutin-related protein	FKRP	4.05046881	9.51612345
UDP-glucuronic acid decarboxylase 1	UXS1	4.05557002	9.48219681
Golgi integral membrane protein 4	GOLIM4	6.2690362	9.46192996
LEM domain-containing protein 2	LEMD2	5.54956912	9.45603085
ATP-dependent zinc metalloprotease YME1L1	YME1L1	5.03557259	9.40953509
Acylglycerol kinase, mitochondrial	AGK	3.66412369	9.40902074
Nuclease EXOG, mitochondrial	EXOG	5.67485247	9.40675958
Polypeptide N-acetylgalactosaminyltransferase 1	GALNT1	2.66075673	9.38523928
Protein O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase 1	POMGNT1	4.88137779	9.37491862
Ras GTPase-activating-like protein IQGAP3	IQGAP3	2.4589743	9.33663527
Nucleolar protein 9	NOP9	3.40668148	9.33001614
Prohibitin-2	PHB2	3.37851277	9.31452147
GPI transamidase component PIG-T	PIGT	2.97560468	9.28877862
Tubulin beta-8 chain	TUBB8	4.67697799	9.24414825
Phosphate carrier protein, mitochondrial	SLC25A3	2.96301928	9.23248227
Glycosylphosphatidylinositol anchor attachment 1 protein	GPAA1	3.46016533	9.20870622
Protein THEM6	THEM6	3.27518293	9.13316631
Melanoma-associated antigen D2	MAGED2	3.02282039	9.1290795
Dehydrogenase/reductase SDR family member 7B	DHRS7B	1.88919553	9.12167931
Uronyl 2-sulfotransferase	UST	4.47026092	9.11200492
Lysophosphatidylcholine acyltransferase 1	LPCAT1	4.75561044	9.10320155
Cleft lip and palate transmembrane protein 1-like protein	CLPTM1L	2.65278714	9.09178797
60S ribosomal protein L38	RPL38	5.6846663	9.08648554
TBC1 domain family member 10B	TBC1D10B	4.84780259	9.07519595
RRP12-like protein	RRP12	3.27635285	9.03566869
Metalloreductase STEAP3	STEAP3	3.21454838	9.02319018
Vesicle-associated membrane protein 8	VAMP8	4.26836341	9.02269681
Tripeptidyl-peptidase 1	TPP1	5.39823267	9.00557073
Exportin-5	XPO5	2.49030681	9.00182311
Peroxisomal membrane protein 2	PXMP2	3.75213584	8.96487013
Vesicle-associated membrane protein 3	VAMP3	3.3557889	8.95338758
Polypeptide N-acetylgalactosaminyltransferase 10	GALNT10	5.02797683	8.94606813
Inositol monophosphatase 3	IMPAD1	5.69324647	8.9263738
Delta(24)-sterol reductase	DHCR24	2.3375547	8.88430214
Exportin-1	XPO1	3.5865221	8.85756397
Rab-like protein 3	RABL3	4.2193174	8.82338842
Syntaxin-4	STX4	5.32423087	8.77082698
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	NDUFA9	3.56550146	8.72957897
Golgi membrane protein 1	GOLM1	4.68134609	8.70034345
Beta-1,4-galactosyltransferase 7;Xylosylprotein 4-beta-galactosyltransferase	B4GALT7	5.03307805	8.69486618
Carbohydrate sulfotransferase 14	CHST14	3.34027584	8.67277463
Phosphorylase b kinase regulatory subunit beta	PHKB	2.20242871	8.66632748
Cytochrome c oxidase subunit 6C	COX6C	4.72101074	8.65423393
NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	NDUFS2	2.40080352	8.63176028
Beta-1,3-galactosyltransferase 6	B3GALT6	4.79877988	8.61066341
Vesicle transport through interaction with t-SNAREs homolog 1A	VTI1A	2.7248909	8.60963472
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	NDUFS3	2.61037458	8.59887664
Fatty acyl-CoA reductase 1	FAR1	2.97315975	8.59555372
Neutral amino acid transporter B(0)	SLC1A5	2.1200329	8.58661747
39S ribosomal protein L11, mitochondrial	MRPL11	4.07169739	8.55678908
Cytochrome c oxidase assembly factor 1 homolog	COA1	4.26827718	8.53657182
Protein phosphatase 1L	PPM1L	4.45605581	8.53190645
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	3.29881213	8.52756914
Fatty aldehyde dehydrogenase	ALDH3A2	5.16354067	8.52295844
Glycoprotein endo-alpha-1,2-mannosidase-like protein	MANEAL	3.87856408	8.48068047
Latrophilin-2	LPHN2	2.65295963	8.46879641
Lysophospholipid acyltransferase LPCAT4	LPCAT4	3.94453663	8.42489147
Cytochrome c oxidase protein 20 homolog	COX20	3.54651992	8.35612233
Peroxisomal membrane protein PMP34	SLC25A17	4.27506248	8.3549188
Peroxisomal membrane protein 11C	PEX11G	4.24172102	8.35473982
Atypical kinase ADCK3, mitochondrial	ADCK3	3.23688681	8.35161304
Soluble calcium-activated nucleotidase 1	CANT1	3.64448292	8.33245405
Selenoprotein H	SELH	3.37499072	8.30301793
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	NDUFB11	3.91752787	8.28105672
Chitobiosyldiphosphodolichol beta-mannosyltransferase	ALG1	3.85054159	8.2588501
Polypeptide N-acetylgalactosaminyltransferase 12	GALNT12	5.46143093	8.23593648
Inactive ubiquitin thioesterase FAM105A	FAM105A	2.85828297	8.22470379
N-acetylglucosamine-1-phosphotransferase subunits alpha/beta	GNPTAB	6.24449569	8.1744744

DnaJ homolog subfamily B member 12	DNAJB12	5.9580305	8.14359951
Nucleolar complex protein 3 homolog	NOC3L	4.73970072	8.14255714
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	4.43413694	8.13474019
[3-methyl-2-oxobutanoate dehydrogenase [lipoamide]] kinase, mitochondrial	BCKDK	2.65282299	8.13108126
Exostosin-2	EXT2	3.64051561	8.11040529
Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	B3GAT3	3.94156271	8.10439777
ATP synthase subunit O, mitochondrial	ATP5O	3.39977166	8.09663741
Tricarboxylate transport protein, mitochondrial	SLC25A1	2.6596078	8.08542156
Transducin beta-like protein 2	TBL2	3.91880477	8.07011445
Thioredoxin-related transmembrane protein 1	TMX1	4.15971473	8.06989225
Ubiquitin carboxyl-terminal hydrolase 22	USP22	2.17306982	8.04077816
Phosphatidylserine decarboxylase proenzyme	PISD	2.74157703	8.00635306
MMS19 nucleotide excision repair protein homolog	MMS19	4.18395332	7.98188591
Cell differentiation protein RCD1 homolog	RQCD1	2.18386334	7.96638648
Xyloside xylosyltransferase 1	XXYLT1	2.69852767	7.96433481
Ras-related protein Ral-A;Ras-related protein Ral-B	RALA;RALB	1.96789763	7.94879023
Peptide deformylase, mitochondrial	PDF	2.12550062	7.94264698
NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	NDUFS1	2.28804284	7.94199149
Importin-4	IPO4	2.2416583	7.9399004
L-lactate dehydrogenase A chain	LDHA	2.63307692	7.89604664
Vesicle-associated membrane protein 4	VAMP4	2.04866115	7.89533234
Aladin	AAAS	1.82013217	7.88804054
Protein CASC4	CASC4	4.41868739	7.87631734
Golgin subfamily B member 1	GOLGB1	3.30711339	7.87377739
Prostaglandin F2 receptor negative regulator	PTGFRN	2.08778546	7.8686622
Glycoprotein endo-alpha-1,2-mannosidase	MANEA	4.38449467	7.81959057
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	HACD3	2.35980116	7.81891696
Protein disulfide-isomerase	P4HB	3.35302508	7.81881936
Probable ergosterol biosynthetic protein 28	C14orf1	5.83346519	7.81390413
ATPase family AAA domain-containing protein 3B	ATAD3B	3.87780896	7.80152321
Receptor-binding cancer antigen expressed on SiSo cells	EBAG9	3.7709393	7.79576715
NLR family member X1	NLRX1	2.40989547	7.78950024
PRA1 family protein 3	ARL6IP5	3.55089616	7.7648414
Polypeptide N-acetylgalactosaminyltransferase 16	GALNT16	2.73469048	7.75729783
60S ribosomal protein L10a	RPL10A	4.83506461	7.74714661
60S ribosomal protein L10	RPL10	3.92424776	7.73391279
Coatomer subunit alpha;Xenin;Proxenin	COPA	2.26090698	7.72839483
Uncharacterized protein C17orf53	C17orf53	3.1414683	7.72780196
Vesicle transport protein GOT1B	GOLT1B	2.3694992	7.71667067
Calnexin	CANX	4.52828376	7.71053759
Immediate early response 3-interacting protein 1	IER3IP1	3.44231829	7.70154413
Trans-Golgi network integral membrane protein 2	TGOLN2	3.03713168	7.68392054
Small integral membrane protein 1	SMIM1	4.01757728	7.682923
Very-long-chain enoyl-CoA reductase	TECR	2.29683118	7.68153222
Coatomer subunit beta	COPB2	2.51019777	7.67975903
Transmembrane protein 33	TMEM33	2.38964967	7.67472839
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	4.56564455	7.63618978
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	NDUFA5	2.35152855	7.62586594
Mitochondrial import inner membrane translocase subunit Tim13	TIMM13	3.55411542	7.6154604
Probable lysosomal cobalamin transporter	LMBRD1	4.65984359	7.61257458
Transmembrane protein 106B	TMEM106B	2.18940635	7.60723623
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B	MGAT4B	4.24904185	7.58674876
Malonyl-CoA-acyl carrier protein transacylase, mitochondrial	MCAT	2.61559649	7.58601316
Mitochondrial import inner membrane translocase subunit TIM50	TIMM50	3.74306822	7.58303801
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	MGAT4A	2.52919665	7.57983939
OCIA domain-containing protein 1	OCIAD1	2.24923782	7.57936255
Cyclin-dependent kinase 1	CDK1	4.71218873	7.49329567
Translocon-associated protein subunit alpha	SSR1	2.8597019	7.49162229
D-glucuronyl C5-epimerase	GLCE	3.08474413	7.47911008
Exostosin-like 3	EXTL3	3.48526983	7.4481074
Coatomer subunit gamma-1	COPG1	2.72519311	7.44248517
B-cell receptor-associated protein 31	BCAP31	3.99527181	7.44115289
Chondroitin sulfate N-acetylgalactosaminyltransferase 2	CSGALNACT2	2.35613058	7.41828823
Protein RER1	RER1	2.76750299	7.36615499
Pituitary tumor-transforming gene 1 protein-interacting protein	PTTG1IP	2.78040265	7.36078707
Integrin beta-1	ITGB1	2.1768817	7.34320227
Cysteine desulfurase, mitochondrial	NFS1	1.54173508	7.32372888
Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	3.13482897	7.30179818
Zinc transporter ZIP14	SLC39A14	3.65590914	7.27716573
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase, mitochondrial	PGS1	2.14645898	7.26635647
Protein FAM162A	FAM162A	2.54979103	7.25662676
Very-long-chain 3-oxoacyl-CoA reductase	HSD17B12	3.11381291	7.25643635
Splicing factor 3B subunit 6	SF3B6	3.16204181	7.23182869
Nurim	NRM	3.18690513	7.22860654
Phosphatidate cytidylyltransferase 2	CDS2	3.0053384	7.22435697
Monocarboxylate transporter 1	SLC16A1	2.60025128	7.20977497
Cytochrome c oxidase assembly protein COX15 homolog	COX15	1.9684544	7.17939536
Syntaxin-8	STX8	3.22066796	7.15394115
WASH complex subunit 7	KIAA1033	2.48831861	7.15225093
Voltage-dependent anion-selective channel protein 1	VDAC1	2.15697326	7.13740985
Nuclear migration protein nudC	NUDC	1.23083053	7.12341245
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IC	MAN1C1	3.67193542	7.12111441
28S ribosomal protein S23, mitochondrial	MRPS23	3.41712206	7.1059742
Transmembrane protein 41B	TMEM41B	1.83990157	7.09772746
Adipocyte plasma membrane-associated protein	APMAP	2.97470365	7.08771324
LYR motif-containing protein 4	LYRM4	2.78892481	7.0745465
60S acidic ribosomal protein P2	RPLP2	2.75164217	7.06719462
Vesicular integral-membrane protein VIP36	LMAN2	3.10310758	7.04494063
Endoplasmic reticulum metallopeptidase 1	ERMP1	1.93139593	6.99037329
Uncharacterized protein KIAA2013	KIAA2013	2.56240003	6.98223909
Beta-1,4-galactosyltransferase 5	B4GALT5	3.66879054	6.96003056
Vitamin K-dependent gamma-carboxylase	GGCX	2.12076409	6.93325456
Single-stranded DNA-binding protein, mitochondrial	SSBP1	4.02196687	6.93128427
2-phosphoxylose phosphatase 1	PXYLP1	3.57891557	6.87355804

Casein kinase I isoform delta	CSNK1D	2.53155903	6.86971378
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	PPP2R5D	1.39087354	6.86936792
ATP-binding cassette sub-family B member 7, mitochondrial	ABCB7	2.39078844	6.86370627
Mitochondrial chaperone BCS1	BCS1L	3.4820105	6.86193307
Protein cornichon homolog 4	CNIH4	3.9359492	6.85324891
L-lactate dehydrogenase B chain	LDHB	2.55212273	6.84636434
Teneurin-3	TENM3	2.45944589	6.8443718
Small integral membrane protein 20	SMIM20	3.4223298	6.8342425
Acyl-CoA desaturase	SCD	5.2959905	6.81335354
Probable 28S rRNA (cytosine-C(5))-methyltransferase	NSUN5	2.22661509	6.80532424
Transcription factor IIIB 90 kDa subunit	BRF1	2.72640674	6.79889202
Membrane-associated progesterone receptor component 1	PGRMC1	2.74612975	6.7860473
ADP/ATP translocase 3;ADP/ATP translocase 3, N-terminally processed	SLC25A6	1.85000474	6.7835385
Membrane protein FAM174A	FAM174A	2.16812785	6.77775224
NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	NDUFS8	2.84505097	6.77302424
Protein LYRIC	MTDH	3.64805968	6.76512082
Evolutionarily conserved signaling intermediate in Toll pathway, mitochondrial	ECSIT	3.85956764	6.76283423
Transmembrane emp24 domain-containing protein 10	TMED10	3.5784697	6.75545375
Ceramide synthase 2	CERS2	2.25605897	6.75410144
Phosphorylase b kinase regulatory subunit alpha, skeletal muscle isoform	PHKA1	2.08053687	6.75115617
Protein kish-A	TMEM167A	2.4483358	6.717302
40S ribosomal protein S12	RPS12	7.03577765	6.69330597
Protein SCO1 homolog, mitochondrial	SCO1	1.99718384	6.67525355
Etoposide-induced protein 2.4 homolog	EI24	2.02584431	6.66191133
5-nucleotidase domain-containing protein 2	NT5DC2	2.35856752	6.62458674
Mono [ADP-ribose] polymerase PARP16	PARP16	1.43501798	6.60988553
Receptor expression-enhancing protein 2	REEP2	3.24474838	6.60833073
Oligosaccharyltransferase complex subunit OSTC	OSTC	2.02752319	6.60719617
G-protein coupled receptor family C group 5 member C	GPRC5C	1.69848231	6.58674367
Nicotinamide phosphoribosyltransferase	NAMPT	3.00685333	6.57203579
Vitamin K epoxide reductase complex subunit 1	VKORC1	2.15612476	6.56542524
Uncharacterized protein C10orf35	C10orf35	1.99580483	6.56406562
Long-chain fatty acid transport protein 4	SLC27A4	1.4768395	6.55614726
Sorting nexin-25	SNX25	2.9134494	6.55241044
Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A	MGAT5	3.02674504	6.53259341
Syntaxin-binding protein 3	STXBP3	1.44353965	6.51800283
Enhancer of rudimentary homolog	ERH	4.27930159	6.50068601
Dihydroorotate dehydrogenase (quinone), mitochondrial	DHODH	1.52458816	6.49859111
TraB domain-containing protein	TRABD	2.37536028	6.48305416
Cytochrome b-c1 complex subunit 9	UQCR10	4.63343396	6.47954464
Integrin alpha-6;Integrin alpha-6 heavy chain	ITGA6	2.20823422	6.44005013
DDRGK domain-containing protein 1	DDRGK1	2.00969894	6.43172487
Mitofusin-1	MFN1	1.46326497	6.42829418
Beta-1,4-galactosyltransferase 4	B4GALT4	2.22589673	6.42418226
Protein FAM134C	FAM134C	2.12361448	6.42242559
Voltage-dependent anion-selective channel protein 2	VDAC2	2.58128665	6.41753387
D-3-phosphoglycerate dehydrogenase	PHGDH	4.131561	6.41333803
Protein-tyrosine sulfotransferase 1	TPST1	2.18050815	6.39738019
Retinol dehydrogenase 14	RDH14	1.63852074	6.39265537
N-acetyltransferase 14	NAT14	2.47231989	6.38550186
Torsin-1A-interacting protein 1	TOR1AIP1	1.3992307	6.3693428
ATP-dependent RNA helicase DDX3X	DDX3X	2.44894548	6.36906719
ADP-ribosylation factor-like protein 10	ARL10	2.61116728	6.36660894
Dephospho-CoA kinase domain-containing protein	DCAKD	1.34014397	6.35096518
ATP synthase subunit gamma, mitochondrial	ATP5C1	2.97759274	6.31724072
Caseinolytic peptidase B protein homolog	CLPB	2.80319737	6.2866443
Coiled-coil domain-containing protein 126	CCDC126	6.41640861	6.26295598
TBC1 domain family member 10A	TBC1D10A	2.09761927	6.2624836
Transmembrane protein 245	TMEM245	2.02401624	6.25396411
Cullin-associated NEDD8-dissociated protein 2	CAND2	2.02184381	6.24390793
Transmembrane 9 superfamily member 4	TM9SF4	1.76475506	6.2089688
Transmembrane protein 35	TMEM35	1.87074532	6.19025771
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 4	NDUFAF4	1.31807891	6.17893759
HIG1 domain family member 2A, mitochondrial	HIGD2A	1.89642177	6.16163985
Chondroitin sulfate glucuronyltransferase	CHPF2	2.92510974	6.16163031
Signal recognition particle receptor subunit beta	SRPRB	1.55113553	6.16103522
Stomatin-like protein 2, mitochondrial	STOML2	3.91345708	6.14081256
Aurora kinase A	AURKA	1.23553711	6.12357457
1-acyl-sn-glycerol-3-phosphate acyltransferase delta	AGPAT4	2.40721453	6.10765489
Desmoplakin	DSP	1.40086355	6.10107581
Protein YIPF3;Protein YIPF3, 36 kDa form III	YIPF3	2.04369696	6.07908948
Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase	ALG12	1.88660342	6.07104429
60S ribosomal protein L36	RPL36	3.51421432	6.06581624
Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	2.22367631	6.06191794
Ubiquinol-cytochrome-c reductase complex assembly factor 2	UQCC2	1.43222604	6.05581983
Lipase maturation factor 2	LMF2	2.23152714	6.04999256
Bystin	BYSL	2.34196566	6.0391372
Tetratricopeptide repeat protein 37	TTC37	1.87904925	6.03670406
Surfeit locus protein 4	SURF4	2.50242761	6.03087076
BRCA1-associated ATM activator 1	BRAT1	3.57786178	6.01761055
CDK5 regulatory subunit-associated protein 2	CDK5RAP2	1.42651694	6.0144132
Xylosyltransferase 2	XYLT2	1.93135328	6.01415634
Serine incorporator 1	SERINC1	2.01258414	6.00423845
Serum paraoxonase/arylesterase 2	PON2	2.33276531	6.00081857
Conserved oligomeric Golgi complex subunit 2	COG2	1.82326481	5.99101734
WW domain-containing oxidoreductase	WWOX	1.93398104	5.98539098
Sorting nexin-14	SNX14	2.49097114	5.98006916
Peroxisomal membrane protein 4	PXMP4	2.9046035	5.97082933
Beta-galactosidase-1-like protein 2	GLB1L2	2.21246226	5.96349621
Serine/threonine-protein kinase 19	STK19	1.63525407	5.95862738
14-3-3 protein gamma;14-3-3 protein gamma, N-terminally processed	YWHAG	2.42059287	5.9444313
Alpha-mannosidase 2x	MAN2A2	2.45138755	5.94162718
Splicing factor 3B subunit 1	SF3B1	2.31331597	5.93033759
40S ribosomal protein S29	RPS29	2.45545333	5.92823919

28S ribosomal protein S28, mitochondrial	MRPS28	2.55539438	5.92664591
UPF0568 protein C14orf166	C14orf166	2.13709661	5.91975753
ATP synthase subunit d, mitochondrial	ATP5H	3.2079757	5.90881952
Mitochondrial amidoxime reducing component 2	MARC2	1.66243287	5.8946565
Carbohydrate sulfotransferase 12	CHST12	2.24275485	5.88986778
Cytochrome P450 2U1	CYP2U1	1.5741433	5.88660685
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 3	NDUFAF3	1.43914162	5.87835058
DnaJ homolog subfamily C member 25	DNAJC25	1.7787504	5.87649028
Myosin light polypeptide 6;Myosin light chain 6B	MYL6/B	2.04166675	5.87463983
Saccharopine dehydrogenase-like oxidoreductase	SCCPDH	1.94978708	5.86877505
Heme oxygenase 1	HMOX1	2.38917295	5.85342216
Leucine-rich repeat and calponin homology domain-containing protein 4	LRCH4	2.01399445	5.85088571
39S ribosomal protein L4, mitochondrial	MRPL4	3.43734958	5.82962354
Ubiquinol-cytochrome-c reductase complex assembly factor 1	UQCC1	1.31168489	5.82716688
Up-regulated during skeletal muscle growth protein 5	USMG5	2.15995447	5.82662106
40S ribosomal protein S30	FAU	1.40178044	5.81603336
Centromere/kinetochore protein zw10 homolog	ZW10	2.84481356	5.80592664
Growth arrest and DNA damage-inducible proteins-interacting protein 1	GADD45GIP1	1.81497079	5.79007848
Translocating chain-associated membrane protein 1	TRAM1	2.01336306	5.78692691
Phosphorylase b kinase gamma catalytic chain, liver/testis isoform	PHKG2	2.14751465	5.78665288
PRA1 family protein 2	PRAF2	2.02905861	5.77247429
Protein-tyrosine sulfotransferase 2	TPST2	1.75255405	5.7578872
Complex I assembly factor TIMMDC1, mitochondrial	TIMMDC1	1.54670347	5.75700283
Transmembrane protein 11, mitochondrial	TMEM11	1.34578085	5.74237378
Phosphatidylglycerophosphatase and protein-tyrosine phosphatase 1	PTPMT1	2.21160368	5.73046748
ATP synthase subunit beta, mitochondrial	ATP5B	4.13161993	5.72570674
Vesicle transport protein SFT2C	SFT2D3	1.9529462	5.72257296
1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	AGPAT1	1.88467873	5.71228313
Insulin receptor substrate 4	IRS4	2.36072704	5.69838301
Vesicle-associated membrane protein-associated protein A	VAPA	2.29981811	5.69605096
Prolactin regulatory element-binding protein	PREB	1.58939903	5.69562213
Splicing factor U2AF 35 kDa subunit	U2AF1	3.34449113	5.67003059
Syntaxin-7	STX7	2.33096134	5.66352971
Unconventional myosin-XIX	MYO19	2.04224792	5.65403144
Emerin	EMD	1.90711612	5.64103762
Cytochrome P450 2S1	CYP2S1	1.52305482	5.63416735
Uncharacterized protein C6orf47	C6orf47	3.32914916	5.62926292
Protein unc-50 homolog	UNC50	1.5515987	5.62414138
Transmembrane protein 246	TMEM246	1.89795472	5.61515935
IQ calmodulin-binding motif-containing protein 1	IQCB1	1.5502443	5.60225423
EH domain-containing protein 1	EHD1	1.33190865	5.59830125
von Willebrand factor A domain-containing protein 8	VWA8	1.63817666	5.56805388
Exonuclease 3-5 domain-containing protein 2	EXD2	2.36039446	5.55796051
Nucleoporin NDC1	NDC1	1.49101737	5.54257075
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	2.27915689	5.54227066
Protein TBRG4	TBRG4	1.95503645	5.54185804
Alpha-1,2-mannosyltransferase ALG9	ALG9	1.58154517	5.53544362
DnaJ homolog subfamily B member 6	DNAJB6	3.13906101	5.52759552
ATP synthase subunit alpha, mitochondrial	ATP5A1	3.57691143	5.52302615
Epidermal growth factor receptor	EGFR	2.55848339	5.51557891
Protein RFT1 homolog	RFT1	1.71240308	5.4986283
Endothelin-converting enzyme 1	ECE1	3.07736365	5.49441783
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A	2.47688488	5.48752912
Protein C3orf33	C3orf33	1.71155841	5.47045453
ATP synthase F(0) complex subunit B1, mitochondrial	ATP5F1	2.2709211	5.46962007
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	NDUFA8	3.01137557	5.45457586
Protein YIF1B	YIF1B	2.19569033	5.4454387
Cleavage and polyadenylation specificity factor subunit 6	CPSF6	1.61672935	5.44444497
UDP-glucose:glycoprotein glucosyltransferase 2	UGGT2	1.56164556	5.4405508
Serine/threonine-protein kinase mTOR	MTOR	1.55660391	5.43652344
Transmembrane emp24 domain-containing protein 1	TMED1	2.71416309	5.43531926
TELO2-interacting protein 2	TTI2	1.36884474	5.42051093
Tubulin beta-4B chain	TUBB4B	4.9013941	5.41905848
14-3-3 protein eta	YWHAH	2.48297881	5.41839155
Beta-1,4-N-acetylgalactosaminyltransferase 3	B4GALNT3	2.37179854	5.39797974
Condensin complex subunit 3	NCAPG	2.01541226	5.39089076
U1 small nuclear ribonucleoprotein C	SNRPC	2.48948648	5.3805205
Reticulon-4	RTN4	3.27706494	5.35218716
Vesicle-associated membrane protein 2	VAMP2	2.40870027	5.33785597
Tubulin beta-2B chain	TUBB2B	2.8471032	5.33307902
Cytochrome b-c1 complex subunit 8	UQCRQ	1.32539358	5.32514032
40S ribosomal protein S5	RPS5	3.54939631	5.30458832
Casein kinase I isoform epsilon	CSNK1E	3.07155735	5.29668013
Transmembrane protein 43	TMEM43	1.51856445	5.29108461
Tubulin alpha-1A chain;Tubulin alpha-3C/D chain	TUBA1A/3C	3.36689452	5.28348859
Long-chain fatty acid transport protein 3	SLC27A3	1.79867984	5.2683169
Ectonucleoside triphosphate diphosphohydrolase 4	ENTPD4	2.01017478	5.25265948
39S ribosomal protein L17, mitochondrial	MRPL17	1.9382584	5.24559307
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	1.27825874	5.24018351
Glycosyltransferase 8 domain-containing protein 1	GLT8D1	1.52387509	5.23756854
Putative ribonuclease	YBEY	1.4084377	5.23471578
Tubulin beta chain	TUBB	5.51580735	5.22996171
TELO2-interacting protein 1 homolog	TTI1	1.68402876	5.21931903
Protein LTV1 homolog	LTV1	1.96810596	5.21414312
Sec1 family domain-containing protein 1	SCFD1	1.38136648	5.20685736
Selenoprotein N	SEPN1	1.61699448	5.20476786
Replication factor C subunit 3	RFC3	2.54768634	5.19536781
Mitochondrial thiamine pyrophosphate carrier	SLC25A19	1.50805845	5.19290765
Acyl-CoA:lysophosphatidylglycerol acyltransferase 1	LPGAT1	1.91777122	5.19212087
Rhomboid-related protein 4	RHBDD1	1.70967692	5.17221006
CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4	ST3GAL4	1.7212871	5.17196623
Exportin-T	XPOT	1.77531528	5.17165025
Reticulocalbin-1	RCN1	2.61298069	5.16183408
Transmembrane protein 161A	TMEM161A	1.870041	5.15716998

Transmembrane and coiled-coil domain-containing protein 1	TMCO1	1.82563018	5.15489546
AarF domain-containing protein kinase 4	ADCK4	2.02221916	5.1538887
Androglobin	ADGB	2.26545341	5.14678446
Coiled-coil domain-containing protein 167	CCDC167	3.32687072	5.14663156
E3 ubiquitin-protein ligase RNF126	RNF126	2.56108147	5.1277914
Transmembrane 9 superfamily member 3	TM9SF3	1.92081557	5.12258307
Long-chain-fatty-acid--CoA ligase 4	ACSL4	2.70481255	5.11245092
Fanconi anemia group A protein	FANCA	2.3730654	5.11213811
Transmembrane 9 superfamily member 1	TM9SF1	2.09244078	5.11060492
Tumor necrosis factor receptor superfamily member 10B	TNFRSF10B	1.82420204	5.09607251
Inactive tyrosine-protein kinase 7	PTK7	2.37466909	5.08936469
Inner nuclear membrane protein Man1	LEMD3	1.69941273	5.08670934
Tubulin-specific chaperone D	TBCD	1.7583702	5.07924716
Histone H2A type 1-C	HIST1H2AC	1.91909658	5.06819916
Integrin beta-6	ITGB6	2.27557423	5.06227748
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	1.60812718	5.05043475
Phosphatidylinositol glycan anchor biosynthesis class U protein	PIGU	2.41825833	5.04009883
HCLS1-associated protein X-1	HAX1	2.5180399	5.02071794
Peroxisomal biogenesis factor 3	PEX3	1.7408082	5.02064228
Uncharacterized protein C1orf43	C1orf43	1.57932669	5.01926899
C-X-C chemokine receptor type 4	CXCR4	1.7147247	5.01289018
Zinc transporter 9	SLC30A9	1.45053813	5.01223596
Glucoside xylosyltransferase 2	GXYLT2	2.52250985	5.00958284
Golgi apparatus membrane protein TVP23 homolog B/C	TVP23B/C	2.23974406	4.99890518
Lamina-associated polypeptide 2, isoforms beta/gamma;Thymopoietin;Thymopentin	TMPO	1.57259671	4.99422805
Nuclear envelope phosphatase-regulatory subunit 1	CNEP1R1	2.04850163	4.98591805
Mediator of RNA polymerase II transcription subunit 23	MED23	1.78531263	4.97756036
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	2.87684525	4.97259394
Tubulin alpha-1B chain;Tubulin alpha-4A chain	TUBA1B/4A	4.28505803	4.97173436
Beta-1,4-galactosyltransferase 3	B4GALT3	2.52137624	4.97108459
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	2.53156034	4.96102778
ATPase family AAA domain-containing protein 1	ATAD1	2.36961137	4.95382245
Neuroplastin	NPTN	1.98918334	4.94562213
Protein MON2 homolog	MON2	1.73686256	4.9453675
Probable ATP-dependent RNA helicase DDX17	DDX17	2.41298221	4.93468761
Non-receptor tyrosine-protein kinase TYK2	TYK2	1.91504645	4.93325488
Uncharacterized protein C15orf61	C15orf61	1.51777581	4.90876007
Symplekin	SYMPK	1.50029889	4.89568774
Golgi pH regulator B;Golgi pH regulator A	GPR89A/B	2.48613181	4.89179961
N-acetylglucosamine-1-phosphotransferase subunit gamma	GNPTG	2.14365974	4.87145837
ATP synthase subunit f, mitochondrial	ATP5J2	1.73216192	4.87043571
Mitotic spindle-associated MMXD complex subunit MIP18	FAM96B	2.45571447	4.86240196
Mitochondrial pyruvate carrier 1	MPC1	1.36262196	4.85794576
Telomere length regulation protein TEL2 homolog	TELO2	1.54673172	4.84533119
Nesprin-3	SYNE3	1.79329464	4.84490204
7-dehydrocholesterol reductase	DHCR7	1.84803912	4.84043694
Thymic stromal cotransporter homolog	SLC46A2	1.75879362	4.83891487
Condensin-2 complex subunit G2	NCAPG2	2.04031729	4.83284664
Pyrroline-5-carboxylate reductase 1, mitochondrial	PYCR1	2.13447465	4.82546202
Dynein assembly factor 5, axonemal	DNAAF5	1.85092404	4.82359187
Probable ATP-dependent RNA helicase DDX5	DDX5	2.28971945	4.80769984
Cell division control protein 42 homolog	CDC42	2.39475325	4.80606016
Diacylglycerol O-acyltransferase 1	DGAT1	2.01668221	4.78072834
Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	1.52909187	4.77097321
40S ribosomal protein S27	RPS27	1.53848863	4.76026853
Nipped-B-like protein	NIPBL	1.63676202	4.75284386
Exportin-6	XPO6	1.51972125	4.75113964
Protein timeless homolog	TIMELESS	1.97419114	4.72938538
Translocon-associated protein subunit gamma	SSR3	2.15391718	4.72697449
Probable methyltransferase TARBP1	TARBP1	2.43004383	4.71626759
Solute carrier family 35 member E1	SLC35E1	1.53171145	4.70601082
Protein SCO2 homolog, mitochondrial	SCO2	1.43219916	4.70381864
Receptor expression-enhancing protein 5	REEP5	1.36356644	4.68161456
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	NDUFB9	2.19508012	4.67901103
Transportin-3	TNPO3	2.28171308	4.67874877
Importin-5	IPO5	2.2728807	4.67580382
60S ribosomal protein L32	RPL32	1.79325054	4.67188613
60S ribosomal protein L29	RPL29	3.63857387	4.66071447
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 2, mitochondrial	NDUFB2	2.07176944	4.66036383
CDGSH iron-sulfur domain-containing protein 2	CISD2	1.58880114	4.65687275
Importin subunit alpha-1	KPNA2	2.53628324	4.63227145
Translocation protein SEC62	SEC62	1.6239351	4.62593365
Diacylglycerol kinase epsilon	DGKE	1.83476946	4.61775335
Protein FAM3A	FAM3A	1.95309782	4.60872968
Plasma membrane calcium-transporting ATPase 1	ATP2B1	1.28327135	4.57554531
LMBR1 domain-containing protein 2	LMBRD2	1.82225963	4.5745395
Spermatid perinuclear RNA-binding protein	STRBP	2.47496742	4.57191149
Glycerol-3-phosphate acyltransferase 4	AGPAT6	1.36033551	4.56372388
Transmembrane emp24 domain-containing protein 9	TMED9	1.48947802	4.5635643
Myosin regulatory light chain 12A;Myosin regulatory light chain 12B	MYL12A/B	1.99870437	4.56237221
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4	NDUFB4	1.96878462	4.54790338
Protein FAM160B1	FAM160B1	1.87823806	4.54572042
HLA class I	HLA-C;HLA-B	2.44305007	4.53036626
Protein AAR2 homolog	AAR2	2.8514856	4.52593486
Condensin-2 complex subunit D3	NCAPD3	1.6897964	4.52572886
Calumenin	CALU	2.48989413	4.51924833
Vitamin K epoxide reductase complex subunit 1-like protein 1	VKORC1L1	2.10496538	4.5149835
Integrin alpha-3;Integrin alpha-3 heavy chain;Integrin alpha-3 light chain	ITGA3	1.75514165	4.50729656
DnaJ homolog subfamily A member 4	DNAJA4	2.8766748	4.50459925
Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase	ALG3	1.96560182	4.50352955
RNA-binding protein 26	RBM26	3.14356264	4.48713239
P2X purinoceptor 4	P2RX4	2.40076589	4.4864734
Endonuclease domain-containing 1 protein	ENDOD1	1.94102899	4.48371951
Ubiquitin	UBC	3.25669985	4.47842058

HIG1 domain family member 1A, mitochondrial	HIGD1A	2.16197353	4.46582222
Tumor necrosis factor receptor superfamily member 10D	TNFRSF10D	2.37285106	4.45759455
Squalene synthase	FDFT1	1.7587591	4.45285447
NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	NDUFV2	2.09647483	4.4452467
Translocon-associated protein subunit delta	SSR4	1.65400185	4.44093482
Methylthioribose-1-phosphate isomerase	MRI1	1.44753784	4.42268054
DNA mismatch repair protein Msh6	MSH6	2.34130029	4.41090806
Signal peptide peptidase-like 2B	SPPL2B	1.83547682	4.40778573
Syntaxin-10	STX10	2.75624475	4.3932085
39S ribosomal protein L12, mitochondrial	MRPL12	2.12389517	4.374451
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform	PPP2R5C	4.59264479	4.3712581
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7	NDUFB7	2.2336683	4.35484346
Importin subunit beta-1	KPNB1	2.79533545	4.34966214
Carboxypeptidase D	CPD	1.5420258	4.33893871
Mitochondrial Rho GTPase 1	RHOT1	1.43063585	4.33852259
Integrator complex subunit 3	INTS3	2.12334713	4.3334322
Vesicle transport protein SFT2B	SFT2D2	2.37834582	4.32035987
Matrix metalloproteinase-15	MMP15	1.67404659	4.31793563
Aurora kinase B	AURKB	1.87097089	4.31662401
Lysine-rich nucleolar protein 1	KNOP1	1.7780608	4.30701447
TATA-box-binding protein;TATA box-binding protein-like protein 2	TBP;TBPL2	3.25726648	4.29675198
28S ribosomal protein S34, mitochondrial	MRPS34	1.84554217	4.27919292
Exostosin-1	EXT1	1.80942199	4.272288
39S ribosomal protein L40, mitochondrial	MRPL40	2.65895368	4.26384894
Zinc finger protein 787	ZNF787	1.65078467	4.26092879
Heparan sulfate N-sulfotransferase 1	NDST1	1.71341572	4.25500425
Mitochondrial glutamate carrier 1;Mitochondrial glutamate carrier 2	SLC25A22/18	1.54075149	4.24775505
ATP-binding cassette sub-family B member 6, mitochondrial	ABCB6	2.03173882	4.23730882
Lysophosphatidic acid receptor 1	LPAR1	1.79186118	4.23465506
Uncharacterized protein C2orf47, mitochondrial	C2orf47	1.92617664	4.22465706
Importin-11	IPO11	2.03618809	4.22431469
Syntaxin-18	STX18	2.42542253	4.21906567
Fukutin	FKTN	1.31283495	4.21870263
14-3-3 protein beta/alpha;14-3-3 protein beta/alpha, N-terminally processed	YWHAB	1.58275115	4.21432336
Protein MMS22-like	MMS22L	2.01993329	4.21243286
Serine/threonine-protein phosphatase 6 regulatory subunit 3	PPP6R3	2.89274224	4.21147315
Guanine nucleotide-binding protein G(z) subunit alpha	GNAZ	3.51947205	4.20165348
Glycosaminoglycan xylosylkinase	FAM20B	2.54585116	4.18721581
ADP-ribosylation factor-like protein 6-interacting protein 6	ARL6IP6	2.4380122	4.18324312
Condensin complex subunit 1	NCAPD2	1.40815323	4.18297863
Fatty acid desaturase 2	FADS2	1.65885006	4.18266964
Sortilin	SORT1	1.76393535	4.17577616
Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase	GCNT1	1.25892295	4.13833936
Lymphoid-specific helicase	HELLS	1.57913648	4.13291995
Neuroguidin	NGDN	2.70513525	4.13030593
Nucleoporin NUP188 homolog	NUP188	1.4673302	4.10679531
Tsukushin	TSKU	2.03684304	4.10290368
E3 ubiquitin-protein ligase listerin	LTN1	1.2913092	4.10255178
ADP/ATP translocase 1	SLC25A4	2.29153306	4.09664027
Trafficking protein particle complex subunit 10	TRAPPC10	1.86187694	4.08972168
Elongation of very long chain fatty acids protein 1	ELOVL1	1.60459493	4.08699417
Neuropathy target esterase	PNPLA6	1.2636398	4.08668709
Integrator complex subunit 1	INTS1	1.42879415	4.08021704
Serpin H1	SERPINH1	2.32950594	4.07698949
Mitochondrial import inner membrane translocase subunit Tim8 A	TIMM8A	2.18739342	4.07075946
Metal transporter CNNM3	CNNM3	1.70105381	4.06066513
Amine oxidase [flavin-containing] A	MAOA	2.49524946	4.05530866
Importin-9	IPO9	2.27351586	4.05385494
Transmembrane emp24 domain-containing protein 4	TMED4	1.52708826	3.99553299
Extended synaptotagmin-1	ESYT1	1.45930486	3.99352296
Folate transporter 1	SLC19A1	1.83399198	3.99018637
Protein unc-45 homolog A	UNC45A	2.53998479	3.98585542
Fanconi anemia group D2 protein	FANCD2	1.35768258	3.98293749
39S ribosomal protein L10, mitochondrial	MRPL10	2.63504244	3.98111629
Nicastrin	NCSTN	3.30185047	3.96842035
Tumor protein p53-inducible protein 11	TP53I11	1.86165196	3.96048133
Tubulin beta-4A chain	TUBB4A	1.46699361	3.94887892
Immunoglobulin superfamily member 8	IGSF8	1.35109724	3.94067128
Coatomer subunit zeta-1	COPZ1	2.23543789	3.93780963
Probable cytosolic iron-sulfur protein assembly protein CIAO1	CIAO1	4.0588042	3.93667507
Cytoplasmic FMR1-interacting protein 1	CYFIP1	1.95781279	3.92997456
Ribonuclease 3	DROSHA	2.27204165	3.88475895
NADH dehydrogenase [ubiquinone] 1 subunit C2	NDUFC2	1.32711232	3.87919553
Calcium and integrin-binding protein 1	CIB1	4.4461991	3.87862364
Small subunit processome component 20 homolog	UTP20	1.98905687	3.87748623
Alpha-mannosidase 2	MAN2A1	1.68260883	3.83557288
40S ribosomal protein S19	RPS19	3.62571866	3.83082326
Elongation of very long chain fatty acids protein 5	ELOVL5	1.5974355	3.81429704
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	NDUFA13	1.88929564	3.80709044
Cholinephosphotransferase 1	CHPT1	2.45059322	3.79482905
Sister chromatid cohesion protein PDS5 homolog A	PDS5A	2.50799455	3.78415998
Nuclear pore complex protein Nup93	NUP93	1.81525826	3.77470843
Angiotensin-converting enzyme;Angiotensin-converting enzyme, soluble form	ACE	1.3241474	3.76583004
Mitochondrial amidoxime-reducing component 1	MARC1	1.27188959	3.74882158
Magnesium transporter protein 1	MAGT1	1.80249222	3.73191516
Excitatory amino acid transporter 1	SLC1A3	2.0073902	3.72594261
Metalloreductase STEAP2	STEAP2	2.5985988	3.71563466
60S ribosomal protein L23	RPL23	3.48989262	3.70565097
LETM1 domain-containing protein LETM2, mitochondrial	LETM2	1.78888966	3.70548407
Dyslexia-associated protein KIAA0319-like protein	KIAA0319L	1.63832824	3.69750182
Nuclear pore complex protein Nup107	NUP107	2.01151658	3.67877865
Dynein light chain roadblock-type 1;Dynein light chain roadblock-type 2	DYNLRB1/2	2.39008461	3.66560713
Adenylate cyclase type 6	ADCY6	2.3077852	3.65684891
Protein YIPF6	YIPF6	2.05833959	3.62226232

Ceramide synthase 1	CERS1	1.36832984	3.60447184
Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1	C1GALT1	1.77081006	3.58750248
40S ribosomal protein S18	RPS18	2.41079106	3.57508755
RNA-binding motif protein, X chromosome	RBMX	2.80896496	3.56926473
Dr1-associated corepressor	DRAP1	1.90122573	3.55137984
SAFB-like transcription modulator	SLTM	3.3956352	3.52605375
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	1.3558621	3.5006307
RuvB-like 1	RUVBL1	2.77338652	3.48382759
Solute carrier family 35 member F6	SLC35F6	1.7967546	3.47736359
Stearoyl-CoA desaturase 5	SCD5	1.65231184	3.46229331
Tubulin gamma-1 chain;Tubulin gamma-2 chain	TUBG1/2	1.82247035	3.4516983
Phosphatidylserine synthase 2	PTDSS2	2.26056104	3.44299984
Heat shock protein 105 kDa	HSPH1	1.47796402	3.40854359
Coatomer subunit delta	ARCN1	1.38571085	3.37481817
Focadhesin	FOCAD	3.18553241	3.36340936
Serine-protein kinase ATM	ATM	1.98272953	3.35469405
Lysosome-associated membrane glycoprotein 2	LAMP2	2.14786339	3.35282453
Sterol regulatory element-binding protein cleavage-activating protein	SCAP	1.90527407	3.34805234
Cytochrome c oxidase subunit 5A, mitochondrial	COX5A	1.38399498	3.33380826
Histone H3.3;Histone H3.1t;Histone H3.1	H3F3A	2.29237693	3.30430921
Sister chromatid cohesion protein PDS5 homolog B	PDS5B	2.07588846	3.28873539
MICOS complex subunit MIC60	IMMT	1.76011837	3.27318414
Mitochondrial fission 1 protein	FIS1	2.82430906	3.26266893
Piezo-type mechanosensitive ion channel component 1	PIEZO1	1.96166758	3.25480493
Golgin subfamily A member 5	GOLGA5	1.97839841	3.24215412
Limb region 1 protein homolog	LMBR1	1.8235793	3.18251101
ATP-dependent RNA helicase DHX36	DHX36	1.83990164	3.16187032
78 kDa glucose-regulated protein	HSPA5	3.24734861	3.04345481
tRNA-splicing ligase RtcB homolog	RTCB	1.38498963	3.0267032
Protein unc-93 homolog B1	UNC93B1	3.35098377	2.97064813
Transmembrane emp24 domain-containing protein 2	TMED2	1.71388224	2.93136374
T-cell leukemia translocation-altered gene protein	TCTA	2.37800138	2.89148235
40S ribosomal protein S3a	RPS3A	2.57031341	2.86236795
40S ribosomal protein S25	RPS25	3.65840015	2.85139179
Solute carrier family 35 member F2	SLC35F2	1.4375683	2.83970642
Cullin-7	CUL7	2.46433943	2.82027245
Heat shock 70 kDa protein 1B;Heat shock 70 kDa protein 1A	HSPA1A/B	2.60967967	2.81980578
RNA-binding protein 27	RBM27	4.90158797	2.74356047
60S ribosomal protein L23a	RPL23A	2.61663454	2.71434689
Centrosomal protein of 192 kDa	CEP192	1.56578258	2.71433544
Splicing factor 3B subunit 2	SF3B2	1.42436316	2.69285425
Signal recognition particle 14 kDa protein	SRP14	1.99021042	2.65150388
MAP7 domain-containing protein 1	MAP7D1	1.94864898	2.56091944
ATP-binding cassette sub-family D member 3	ABCD3	1.64519845	2.5107193
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	2.3680206	2.48908901
Zinc finger CCCH domain-containing protein 4	ZC3H4	1.31607853	2.40971724
Small nuclear ribonucleoprotein Sm D2	SNRPD2	2.0500359	2.38885657
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	2.13468154	2.36450863
G-rich sequence factor 1	GRSF1	1.4052482	2.31181304
DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H	3.23809799	2.30574258
Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	1.75727839	2.25640011
Y-box-binding protein 3	YBX3	1.56680568	2.01809025
40S ribosomal protein S16	RPS16	2.21502678	1.94703611
Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	1.31414563	1.7140824
Nesprin-2	SYNE2	2.03296932	1.64078776
40S ribosomal protein SA	RPSA	1.92433342	1.50723807
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	1.5467159	1.47701581

Supplementary Table 8.6 GST-GOLPH3 vs GST-GOLPH3L

Protein name	Gene name	-Log$_{10}$ p-value	Difference Log$_2$(GOLPH3/ GOLPH3L)
Myosin-10	MYH10	2.86797601	9.83817673
Mitofusin-1	MFN1	3.34723929	9.18662834
Metallo-beta-lactamase domain-containing protein 2	MBLAC2	3.93822779	8.62978427
Transitional endoplasmic reticulum ATPase	VCP	1.97090306	8.4445591
Aurora kinase A	AURKA	3.23920393	8.26705933
Dephospho-CoA kinase domain-containing protein	DCAKD	3.04633641	8.26577314
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 4	NDUFAF4	3.32710904	8.12132645
NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial	NDUFS7	3.51340914	8.10161463
Mono [ADP-ribose] polymerase PARP16	PARP16	3.32690959	8.08737628
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	PPP2R5D	1.37461086	7.65182495
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 3	NDUFAF3	2.84116482	7.23985863
E3 UFM1-protein ligase 1	UFL1	1.85796932	7.14205678
Torsin-1A-interacting protein 1	TOR1AIP1	2.74484114	7.02665138
Glutathione synthetase	GSS	2.71886965	6.95416196
Retinol dehydrogenase 14	RDH14	1.84426477	6.72413317
Putative ribonuclease	YBEY	2.24813936	6.7050546
CDK5 regulatory subunit-associated protein 2	CDK5RAP2	2.41316863	6.59240786

Casein kinase I isoform delta	CSNK1D	1.93739643	6.57248624
Serine/threonine-protein kinase 19	STK19	1.94361663	6.45858637
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	2.34041635	6.3831323
DNA repair protein RAD50	RAD50	1.88899204	6.33711306
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	NDUFA5	1.77863465	6.26917775
Signal recognition particle receptor subunit beta	SRPRB	2.25169916	6.14471372
Growth arrest and DNA damage-inducible proteins-interacting protein 1	GADD45GIP1	2.29101732	6.03960228
Protein SCO1 homolog, mitochondrial	SCO1	1.89869173	5.96706835
Coiled-coil-helix-coiled-coil-helix domain-containing protein 1	CHCHD1	2.91539684	5.9587663
Ras-related protein Ral-A;Ras-related protein Ral-B	RALA;RALB	1.40036412	5.91065025
NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	NDUFS2	2.59617083	5.88941956
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	NDUFS3	1.49105894	5.7642765
Conserved oligomeric Golgi complex subunit 2	COG2	2.96774854	5.62756602
Aurora kinase B	AURKB	2.88591117	5.41756058
Inner nuclear membrane protein Man1	LEMD3	3.04562121	5.38691902
Transcription factor HES-4	HES4	2.16348194	5.33312798
Heme oxygenase 1	HMOX1	2.84460264	5.27674421
Tumor necrosis factor receptor superfamily member 10B	TNFRSF10B	2.62250404	5.22025363
Long-chain fatty acid transport protein 3	SLC27A3	2.3262568	5.1465683
Protein timeless homolog	TIMELESS	2.42831166	4.99819311
Telomere length regulation protein TEL2 homolog	TELO2	1.63335725	4.99286016
NADH dehydrogenase [ubiquinone] 1 subunit C2	NDUFC2;	2.527495	4.86882019
Angiotensin-converting enzyme;Angiotensin-converting enzyme, soluble form	ACE	1.45755927	4.80877177
Uncharacterized protein C1orf43	C1orf43	2.210192	4.73214404
GH3 domain-containing protein	GHDC	1.89742526	4.6924146
Translocon-associated protein subunit delta	SSR4	2.91692289	4.65713692
Protein FAM160B1	FAM160B1	1.70306967	4.63937823
Protein TBRG4	TBRG4	1.39721623	4.61416245
Lamina-associated polypeptide 2, isoforms beta/gamma;Thymopoietin;Thymopentin	TMPO	1.51024443	4.56173261
Mitochondrial amidoxime-reducing component 1	MARC1	1.80609739	4.52524312
Vesicle-associated membrane protein 2	VAMP2	2.26585414	4.38420614
14-3-3 protein eta	YWHAH	2.50404737	4.38308589
Protein SCO2 homolog, mitochondrial	SCO2	1.77030368	4.34916751
Nipped-B-like protein	NIPBL	1.61530494	4.32196744
Uncharacterized protein C15orf61	C15orf61	1.77998125	4.30416107
Vesicle-associated membrane protein-associated protein A	VAPA	1.64095694	4.10742442
Golgi phosphoprotein 3	GOLPH3	2.77881005	4.07718404
Ubiquitin carboxyl-terminal hydrolase 7	USP7	1.89204053	4.04847844
ATP-dependent RNA helicase DDX3X;ATP-dependent RNA helicase DDX3Y	DDX3X;DDX3Y	2.50294151	3.85914675
Membrane-associated progesterone receptor component 2	PGRMC2	1.44802366	3.84567324
NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	NDUFS8	2.28502473	3.82557424
Atypical kinase ADCK3, mitochondrial	ADCK3	1.88488877	3.76907921
Protein THEM6	THEM6	1.43605096	3.70592054
Nucleolar protein 9	NOP9	1.31262225	3.64964612
Vesicle-associated membrane protein 3	VAMP3	1.54087723	3.55036799
ATP-dependent RNA helicase DDX1	DDX1	1.59973051	3.54773521
Integrator complex subunit 3	INTS3	2.41437585	3.54704793
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	NDUFA13	1.71352471	3.50098674
Transcription factor IIIB 90 kDa subunit	BRF1	1.50656916	3.49508349
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4	NDUFB4	1.80446793	3.20783615
Myosin light polypeptide 6;Myosin light chain 6B	MYL6;MYL6B	1.98807793	3.12335332
Mitochondrial import inner membrane translocase subunit Tim13	TIMM13	1.51818977	2.94052315
Replication factor C subunit 3	RFC3	1.39000853	2.8930556
DnaJ homolog subfamily C member 10	DNAJC10	1.33662202	2.87614123
Golgin subfamily A member 5	GOLGA5	1.65050242	2.82916005
Probable ATP-dependent RNA helicase DDX5	DDX5	1.7679032	2.79029592
Histone H2A.V;Histone H2A.Z	H2AFV;H2AFZ	1.94142793	2.77122116
LETM1 and EF-hand domain-containing protein 1, mitochondrial	LETM1	1.68474213	2.71751531
Protein CASP	CUX1	2.00917537	2.24006144
Nucleolar complex protein 3 homolog	NOC3L	1.90234555	1.88439687
Ubiquitin-like modifier-activating enzyme 1	UBA1	1.71097867	1.70132701
26S proteasome non-ATPase regulatory subunit 1	PSMD1	1.36555942	1.6396122
Calnexin	CANX	1.56537194	1.2883962

Glutathione S-transferase C-terminal domain-containing protein	GSTCD	1.73178981	1.26317406
Poly [ADP-ribose] polymerase 1	PARP1	1.55650634	1.14027659
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	PPP2R1A	1.44007693	1.10214742
Vimentin	VIM	2.40380445	-0.6684895
Asparagine--tRNA ligase, cytoplasmic	NARS	2.06512786	-0.6786207
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	1.41738372	-1.0342445
Probable asparagine--tRNA ligase, mitochondrial	NARS2	1.5246927	-1.1774216
Interleukin enhancer-binding factor 2	ILF2	2.13837234	-1.273948
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	1.58215367	-1.3149796
Glutaminyl-peptide cyclotransferase-like protein	QPCTL	1.66496011	-1.3182456
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	1.3896528	-1.3380896
Golgi integral membrane protein 4	GOLIM4	1.67572847	-1.4150899
Nuclease-sensitive element-binding protein 1	YBX1	1.47933072	-1.4180342
60S ribosomal protein L15	RPL15	1.43739461	-1.4295979
T-complex protein 1 subunit alpha	TCP1	1.92226765	-1.4716117
Valine--tRNA ligase	VARS	1.73117607	-1.4720713
Protein FAM3C	FAM3C	1.32172088	-1.4741974
SAFB-like transcription modulator	SLTM	1.82168131	-1.6474609
Splicing factor, proline- and glutamine-rich	SFPQ	1.65425698	-1.9771945
39S ribosomal protein L40, mitochondrial	MRPL40	1.52063006	-2.0699437
Nuclear export mediator factor NEMF	NEMF	2.29520201	-2.1235555
G-rich sequence factor 1	GRSF1	1.3069511	-2.1632442
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	1.36947005	-2.2625777
Uronyl 2-sulfotransferase	UST	1.32498123	-2.4186624
Histone H1.2	HIST1H1C	2.05556416	-2.4377073
UDP-glucuronic acid decarboxylase 1	UXS1	1.31897356	-2.5595614
SRSF protein kinase 1	SRPK1	1.54218394	-2.6661371
Zinc finger RNA-binding protein	ZFR	1.5196013	-2.6691767
Glycosaminoglycan xylosylkinase	FAM20B	2.10038391	-2.687802
Protein disulfide-isomerase	P4HB	1.83619286	-2.7151241
ESF1 homolog	ESF1	1.39195688	-2.7631124
RNA-binding protein 39	RBM39	1.36132252	-2.7715092
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	1.45419578	-2.7804616
39S ribosomal protein L11, mitochondrial	MRPL11	2.01095662	-2.9347388
Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	B3GAT3	2.79959896	-2.9666246
Nucleosome assembly protein 1-like 1	NAP1L1	2.03759064	-2.9818694
Vesicle transport protein SFT2B	SFT2D2	1.31376486	-3.2687213
ATP-dependent RNA helicase DDX50	DDX50	1.87941171	-3.3058561
Folate transporter 1	SLC19A1	1.42173568	-3.336661
Double-stranded RNA-binding protein Staufen homolog 1	STAU1	1.48629513	-3.6536681
Treacle protein	TCOF1	1.84437917	-3.7281017
ADP-ribosylation factor-like protein 10	ARL10	1.45679162	-3.7713642
Condensin complex subunit 3	NCAPG	1.95725946	-3.8260244
Signal recognition particle 9 kDa protein	SRP9	1.31074859	-3.9173571
N-acetyltransferase 10	NAT10	1.96495266	-3.9687169
Non-POU domain-containing octamer-binding protein	NONO	1.60549508	-3.9791018
60S ribosomal protein L19	RPL19	4.64903167	-4.1278083
La-related protein 1	LARP1	1.73863318	-4.1482099
tRNA (guanine(37)-N1)-methyltransferase	TRMT5	2.29388702	-4.2086957
Glutamyl-tRNA(Gln) amidotransferase subunit C, mitochondrial	GATC	3.24241935	-4.5653254
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	2.84785003	-4.5661392
Nucleosome assembly protein 1-like 4	NAP1L4	1.43805949	-4.5878798
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	3.18089419	-4.7630965
Ribosomal L1 domain-containing protein 1	RSL1D1	1.60445746	-4.8374055
Probable ATP-dependent RNA helicase DDX41	DDX41	2.95311695	-4.9735305
CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4	ST3GAL4	3.02983844	-5.1002146
Serine incorporator 1	SERINC1	1.87435543	-5.1295268
ATP-dependent RNA helicase DDX18	DDX18	1.90111695	-5.3217322
Importin subunit alpha-4	KPNA3	3.73482455	-5.3877207
40S ribosomal protein S21	RPS21	2.50672938	-5.3976212
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	3.39458725	-5.4337711
rRNA methyltransferase 3, mitochondrial	RNMTL1	1.41806525	-5.505223
Fukutin	FKTN	2.00454108	-5.5425625
Probable ATP-dependent RNA helicase DDX47	DDX47	2.82506475	-5.616347
Protein-tyrosine sulfotransferase 2	TPST2	2.08958667	-5.6603495
Glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial	GATB	2.04095372	-5.8156586
Glutamyl-tRNA(Gln) amidotransferase subunit A, mitochondrial	QRSL1	2.40929161	-6.5599194
EH domain-containing protein 1	EHD1	2.45412293	-6.5911547
WW domain-containing oxidoreductase	WWOX	3.27644789	-7.3386955

Ubiquinol-cytochrome-c reductase complex assembly factor 2	UQCC2	1.72573384	-7.3990714
Nuclear migration protein nudC	NUDC	3.3721216	-9.7550335
Golgi phosphoprotein 3-like	GOLPH3L	3.40218952	-14.953658

Supplementary Table 8.7 COPI proteome, GOLPH3+3L and FAM114A2 interactome Venn diagram

Gene Name	Grouping
ACSL3	COPI Proteome
ACTB	COPI Proteome
ACTN4	COPI Proteome
ADAM10	COPI Proteome
ADAM9	COPI Proteome
AGTRAP	COPI Proteome
AP1G1	COPI Proteome
AP2A1	COPI Proteome
AP2B1	COPI Proteome
ARFRP1	COPI Proteome
ATG9A	COPI Proteome
ATL2	COPI Proteome
ATL3	COPI Proteome
ATP1B1	COPI Proteome
ATP2C1	COPI Proteome
ATP6AP1	COPI Proteome
ATP6V0A1	COPI Proteome
ATP6V0A2	COPI Proteome
ATP6V0D1	COPI Proteome
ATP7A	COPI Proteome
AXL	COPI Proteome
B4GALT1	COPI Proteome
BET1	COPI Proteome
C1GALT1C1	COPI Proteome
C7orf50	COPI Proteome
CALR	COPI Proteome
CAV2	COPI Proteome
CD70	COPI Proteome
CGREF1	COPI Proteome
CLINT1	COPI Proteome
CLPTM1	COPI Proteome
CLU	COPI Proteome
CTNNA1	COPI Proteome
CTNNB1	COPI Proteome
CTSD	COPI Proteome
CYB5R3	COPI Proteome
DNAJC13	COPI Proteome
DSE	COPI Proteome
EDIL3	COPI Proteome
EPHX1	COPI Proteome
ERGIC1	COPI Proteome
ERGIC2	COPI Proteome
ERGIC3	COPI Proteome
ERP44	COPI Proteome
FAM177A1	COPI Proteome
GANAB	COPI Proteome
GNAI1	COPI Proteome
GNB4	COPI Proteome
GOLGA2	COPI Proteome
GOLPH3	COPI Proteome
GOSR2	COPI Proteome
GPR107	COPI Proteome
GPR108	COPI Proteome
GPX8	COPI Proteome
HLA-A	COPI Proteome
HM13	COPI Proteome
HNRNPC	COPI Proteome
HNRPM	COPI Proteome
IFITM1	COPI Proteome
IGF1R	COPI Proteome
IGFBP7	COPI Proteome
ITGA2	COPI Proteome
ITGB5	COPI Proteome
KCT2	COPI Proteome
KDELR1	COPI Proteome
L1CAM	COPI Proteome
LDLR	COPI Proteome
LGALS3BP	COPI Proteome
LMAN1	COPI Proteome
LMAN2L	COPI Proteome
LNPEP	COPI Proteome
LPL	COPI Proteome
LSS	COPI Proteome
MGAT1	COPI Proteome
MYADM	COPI Proteome
MYH10	COPI Proteome
MYH9	COPI Proteome
MYO1E	COPI Proteome
MYOF	COPI Proteome
NCL	COPI Proteome
NPM1	COPI Proteome
NSDHL	COPI Proteome

NUCB1	COPI Proteome
NUCB2	COPI Proteome
PDIA3	COPI Proteome
PDIA6	COPI Proteome
PGK1	COPI Proteome
PGRMC2	COPI Proteome
PI4K2A	COPI Proteome
POR	COPI Proteome
PPIB	COPI Proteome
PRDX1	COPI Proteome
PSAP	COPI Proteome
RAB10	COPI Proteome
RAB11B	COPI Proteome
RAB14	COPI Proteome
RAB18	COPI Proteome
RAB1A	COPI Proteome
RAB1B	COPI Proteome
RAB21	COPI Proteome
RAB2A	COPI Proteome
RAB32	COPI Proteome
RAB5A	COPI Proteome
RAB5B	COPI Proteome
RAB5C	COPI Proteome
RAB6A	COPI Proteome
RAB7A	COPI Proteome
RABAC1	COPI Proteome
ROBO1	COPI Proteome
RTN3	COPI Proteome
SCAMP3	COPI Proteome
SCAMP4	COPI Proteome
SCARB1	COPI Proteome
SDF4	COPI Proteome
SF3B3	COPI Proteome
SGMS2	COPI Proteome
SLC12A4	COPI Proteome
SLC1A4	COPI Proteome
SLC30A5	COPI Proteome
SLC30A6	COPI Proteome
SLC30A7	COPI Proteome
SLC35A2	COPI Proteome
SLC35B2	COPI Proteome
SLC35C1	COPI Proteome
SLC35E2	COPI Proteome
SLC38A2	COPI Proteome
SLC43A3	COPI Proteome
SMOC1	COPI Proteome
SNAP23	COPI Proteome
STRF8	COPI Proteome
STX5	COPI Proteome
TM9SF2	COPI Proteome
TMED3	COPI Proteome
TMED5	COPI Proteome
TMED7	COPI Proteome
TMEM115	COPI Proteome
TMEM165	COPI Proteome
TMEM168	COPI Proteome
TMEM23	COPI Proteome
TMEM87A	COPI Proteome
TPBG	COPI Proteome
TRA1	COPI Proteome
UBIAD1	COPI Proteome
YIF1A	COPI Proteome
YIPF4	COPI Proteome
YIPF5	COPI Proteome
ZDHHC13	COPI Proteome
AAR2	GOLPH3+3L Interactor
ABCB10	GOLPH3+3L Interactor
ABCB6	GOLPH3+3L Interactor
ABCB7	GOLPH3+3L Interactor
ABHD12	GOLPH3+3L Interactor
ACAD9	GOLPH3+3L Interactor
ACE	GOLPH3+3L Interactor
ACSL4	GOLPH3+3L Interactor
ADCK3	GOLPH3+3L Interactor
ADCK4	GOLPH3+3L Interactor
ADCY6	GOLPH3+3L Interactor
ADGB	GOLPH3+3L Interactor
AFG3L2	GOLPH3+3L Interactor
AGK	GOLPH3+3L Interactor
AGPAT1	GOLPH3+3L Interactor
AGPAT4	GOLPH3+3L Interactor
AGPAT6	GOLPH3+3L Interactor
ALG1	GOLPH3+3L Interactor
ALG12	GOLPH3+3L Interactor
ALG3	GOLPH3+3L Interactor
ALG9	GOLPH3+3L Interactor
ARCN1	GOLPH3+3L Interactor
ARL10	GOLPH3+3L Interactor
ARL6IP1	GOLPH3+3L Interactor
ARL6IP6	GOLPH3+3L Interactor
ARMCX3	GOLPH3+3L Interactor
ATAD1	GOLPH3+3L Interactor
ATAD3A	GOLPH3+3L Interactor
ATAD3B	GOLPH3+3L Interactor

ATM	GOLPH3+3L Interactor
ATP5A1	GOLPH3+3L Interactor
ATP5B	GOLPH3+3L Interactor
ATP5F1	GOLPH3+3L Interactor
ATP5H	GOLPH3+3L Interactor
ATP5J2	GOLPH3+3L Interactor
ATP5O	GOLPH3+3L Interactor
AURKA	GOLPH3+3L Interactor
AURKB	GOLPH3+3L Interactor
B3GALT6	GOLPH3+3L Interactor
B4GALNT3	GOLPH3+3L Interactor
B4GALT3	GOLPH3+3L Interactor
B4GALT4	GOLPH3+3L Interactor
B4GALT7	GOLPH3+3L Interactor
BCKDK	GOLPH3+3L Interactor
BCS1L	GOLPH3+3L Interactor
BRAT1	GOLPH3+3L Interactor
BRF1	GOLPH3+3L Interactor
BRI3BP	GOLPH3+3L Interactor
BYSL	GOLPH3+3L Interactor
C10orf35	GOLPH3+3L Interactor
C14orf1	GOLPH3+3L Interactor
C14orf166	GOLPH3+3L Interactor
C15orf61	GOLPH3+3L Interactor
C17orf53	GOLPH3+3L Interactor
C1orf43	GOLPH3+3L Interactor
C2orf47	GOLPH3+3L Interactor
C3orf33	GOLPH3+3L Interactor
C6orf47	GOLPH3+3L Interactor
CAND2	GOLPH3+3L Interactor
CCDC126	GOLPH3+3L Interactor
CCDC167	GOLPH3+3L Interactor
CDC42	GOLPH3+3L Interactor
CDK1	GOLPH3+3L Interactor
CDK5RAP2	GOLPH3+3L Interactor
CDS2	GOLPH3+3L Interactor
CEBPZ	GOLPH3+3L Interactor
CEP192	GOLPH3+3L Interactor
CERS1	GOLPH3+3L Interactor
CERS2	GOLPH3+3L Interactor
CHP1	GOLPH3+3L Interactor
CHPT1	GOLPH3+3L Interactor
CHST12	GOLPH3+3L Interactor
CIAO1	GOLPH3+3L Interactor
CIB1	GOLPH3+3L Interactor
CISD2	GOLPH3+3L Interactor
CLPB	GOLPH3+3L Interactor
CNEP1R1	GOLPH3+3L Interactor
CNNM3	GOLPH3+3L Interactor
COA1	GOLPH3+3L Interactor
COG2	GOLPH3+3L Interactor
COPA	GOLPH3+3L Interactor
COPB1	GOLPH3+3L Interactor
COPB2	GOLPH3+3L Interactor
COPE	GOLPH3+3L Interactor
COPG1	GOLPH3+3L Interactor
COPG2	GOLPH3+3L Interactor
COPZ1	GOLPH3+3L Interactor
COX15	GOLPH3+3L Interactor
COX20	GOLPH3+3L Interactor
COX5A	GOLPH3+3L Interactor
COX6C	GOLPH3+3L Interactor
CPSF6	GOLPH3+3L Interactor
CPT1A	GOLPH3+3L Interactor
CSGALNACT2	GOLPH3+3L Interactor
CSNK1D	GOLPH3+3L Interactor
CSNK1E	GOLPH3+3L Interactor
CUL7	GOLPH3+3L Interactor
CXCR4	GOLPH3+3L Interactor
CYFIP1	GOLPH3+3L Interactor
CYP2S1	GOLPH3+3L Interactor
CYP2U1	GOLPH3+3L Interactor
DARS2	GOLPH3+3L Interactor
DCAKD	GOLPH3+3L Interactor
DDRGK1	GOLPH3+3L Interactor
DDX17	GOLPH3+3L Interactor
DDX3X;DDX3Y	GOLPH3+3L Interactor
DDX5	GOLPH3+3L Interactor
DGAT1	GOLPH3+3L Interactor
DGKE	GOLPH3+3L Interactor
DHCR7	GOLPH3+3L Interactor
DHODH	GOLPH3+3L Interactor
DHRS13	GOLPH3+3L Interactor
DHRS7B	GOLPH3+3L Interactor
DHX36	GOLPH3+3L Interactor
DNAAF5	GOLPH3+3L Interactor
DNAJA4	GOLPH3+3L Interactor
DNAJB12	GOLPH3+3L Interactor
DNAJB6	GOLPH3+3L Interactor
DNAJC19	GOLPH3+3L Interactor
DNAJC25	GOLPH3+3L Interactor
DPM1	GOLPH3+3L Interactor
DRAP1	GOLPH3+3L Interactor
DROSHA	GOLPH3+3L Interactor

DSP	GOLPH3+3L Interactor
DYNLRB1;DYNLRB2	GOLPH3+3L Interactor
EBAG9	GOLPH3+3L Interactor
ECSIT	GOLPH3+3L Interactor
EFNB1	GOLPH3+3L Interactor
EHD1	GOLPH3+3L Interactor
EI24	GOLPH3+3L Interactor
EIF2AK2	GOLPH3+3L Interactor
ELOVL1	GOLPH3+3L Interactor
ELOVL5	GOLPH3+3L Interactor
ENDOD1	GOLPH3+3L Interactor
ENTPD4	GOLPH3+3L Interactor
ERH	GOLPH3+3L Interactor
ERMP1	GOLPH3+3L Interactor
EXD2	GOLPH3+3L Interactor
EXT1	GOLPH3+3L Interactor
EXT2	GOLPH3+3L Interactor
EXTL3	GOLPH3+3L Interactor
FADS2	GOLPH3+3L Interactor
FAF2	GOLPH3+3L Interactor
FAM105A	GOLPH3+3L Interactor
FAM134C	GOLPH3+3L Interactor
FAM160B1	GOLPH3+3L Interactor
FAM162A	GOLPH3+3L Interactor
FAM174A	GOLPH3+3L Interactor
FAM3A	GOLPH3+3L Interactor
FAM96B	GOLPH3+3L Interactor
FANCA	GOLPH3+3L Interactor
FANCD2	GOLPH3+3L Interactor
FAR1	GOLPH3+3L Interactor
FAU	GOLPH3+3L Interactor
FDFT1	GOLPH3+3L Interactor
FIS1	GOLPH3+3L Interactor
FKRP	GOLPH3+3L Interactor
FKTN	GOLPH3+3L Interactor
FOCAD	GOLPH3+3L Interactor
GADD45GIP1	GOLPH3+3L Interactor
GALNT16	GOLPH3+3L Interactor
GALNT6	GOLPH3+3L Interactor
GCN1L1	GOLPH3+3L Interactor
GCNT1	GOLPH3+3L Interactor
GDAP1	GOLPH3+3L Interactor
GGCX	GOLPH3+3L Interactor
GNAZ	GOLPH3+3L Interactor
GNPTAB	GOLPH3+3L Interactor
GNPTG	GOLPH3+3L Interactor
GOLGA5	GOLPH3+3L Interactor
GPAA1	GOLPH3+3L Interactor
GPR89B;GPR89A	GOLPH3+3L Interactor
GPRC5C	GOLPH3+3L Interactor
GRSF1	GOLPH3+3L Interactor
GXYLT1	GOLPH3+3L Interactor
GXYLT2	GOLPH3+3L Interactor
H3F3A	GOLPH3+3L Interactor
HACD3	GOLPH3+3L Interactor
HAX1	GOLPH3+3L Interactor
HELLS	GOLPH3+3L Interactor
HIGD2A	GOLPH3+3L Interactor
HIST1H2AC	GOLPH3+3L Interactor
HLA-C;HLA-B	GOLPH3+3L Interactor
HNRNPAB	GOLPH3+3L Interactor
HNRNPD	GOLPH3+3L Interactor
HNRNPH1	GOLPH3+3L Interactor
HSD17B12	GOLPH3+3L Interactor
HSPA1B;HSPA1A	GOLPH3+3L Interactor
HSPH1	GOLPH3+3L Interactor
IGF2BP1	GOLPH3+3L Interactor
IGSF8	GOLPH3+3L Interactor
INTS1	GOLPH3+3L Interactor
INTS3	GOLPH3+3L Interactor
IPO11	GOLPH3+3L Interactor
IPO4	GOLPH3+3L Interactor
IPO5	GOLPH3+3L Interactor
IPO9	GOLPH3+3L Interactor
IQCB1	GOLPH3+3L Interactor
IQGAP3	GOLPH3+3L Interactor
IRS4	GOLPH3+3L Interactor
ITGA3	GOLPH3+3L Interactor
ITGA6	GOLPH3+3L Interactor
ITGB1	GOLPH3+3L Interactor
ITGB6	GOLPH3+3L Interactor
KIAA1033	GOLPH3+3L Interactor
KNOP1	GOLPH3+3L Interactor
KPNA2	GOLPH3+3L Interactor
KPNB1	GOLPH3+3L Interactor
LAMP2	GOLPH3+3L Interactor
LDHA	GOLPH3+3L Interactor
LEMD3	GOLPH3+3L Interactor
LETM1	GOLPH3+3L Interactor
LETM2	GOLPH3+3L Interactor
LETMD1	GOLPH3+3L Interactor
LMBR1	GOLPH3+3L Interactor
LMBRD1	GOLPH3+3L Interactor
LMF2	GOLPH3+3L Interactor

LPAR1	GOLPH3+3L Interactor
LPCAT1	GOLPH3+3L Interactor
LPCAT4	GOLPH3+3L Interactor
LPGAT1	GOLPH3+3L Interactor
LPHN2	GOLPH3+3L Interactor
LRCH4	GOLPH3+3L Interactor
LRRC59	GOLPH3+3L Interactor
LTN1	GOLPH3+3L Interactor
LTV1	GOLPH3+3L Interactor
LYRM4	GOLPH3+3L Interactor
MAGED2	GOLPH3+3L Interactor
MAGT1	GOLPH3+3L Interactor
MAN1C1	GOLPH3+3L Interactor
MAN2A2	GOLPH3+3L Interactor
MANEA	GOLPH3+3L Interactor
MAOA	GOLPH3+3L Interactor
MAP7D1	GOLPH3+3L Interactor
MARC1	GOLPH3+3L Interactor
MARC2	GOLPH3+3L Interactor
MCAT	GOLPH3+3L Interactor
MED23	GOLPH3+3L Interactor
MFN1	GOLPH3+3L Interactor
MGAT4A	GOLPH3+3L Interactor
MGAT5	GOLPH3+3L Interactor
MMP15	GOLPH3+3L Interactor
MMS19	GOLPH3+3L Interactor
MMS22L	GOLPH3+3L Interactor
MON2	GOLPH3+3L Interactor
MPC1	GOLPH3+3L Interactor
MRI1	GOLPH3+3L Interactor
MRPL10	GOLPH3+3L Interactor
MRPL11	GOLPH3+3L Interactor
MRPL12	GOLPH3+3L Interactor
MRPL17	GOLPH3+3L Interactor
MRPL4	GOLPH3+3L Interactor
MRPL40	GOLPH3+3L Interactor
MRPS23	GOLPH3+3L Interactor
MRPS28	GOLPH3+3L Interactor
MRPS34	GOLPH3+3L Interactor
MSH6	GOLPH3+3L Interactor
MTDH	GOLPH3+3L Interactor
MTOR	GOLPH3+3L Interactor
MYBBP1A	GOLPH3+3L Interactor
MYL12A;MYL12B	GOLPH3+3L Interactor
MYL6;MYL6B	GOLPH3+3L Interactor
MYO19	GOLPH3+3L Interactor
NAMPT	GOLPH3+3L Interactor
NAT14	GOLPH3+3L Interactor
NCAPD2	GOLPH3+3L Interactor
NCAPD3	GOLPH3+3L Interactor
NCAPG	GOLPH3+3L Interactor
NCAPG2	GOLPH3+3L Interactor
NCBP2-AS2	GOLPH3+3L Interactor
NDC1	GOLPH3+3L Interactor
NDST1	GOLPH3+3L Interactor
NDUFA13	GOLPH3+3L Interactor
NDUFA4	GOLPH3+3L Interactor
NDUFA5	GOLPH3+3L Interactor
NDUFA8	GOLPH3+3L Interactor
NDUFA9	GOLPH3+3L Interactor
NDUFAF3	GOLPH3+3L Interactor
NDUFAF4	GOLPH3+3L Interactor
NDUFB10	GOLPH3+3L Interactor
NDUFB11	GOLPH3+3L Interactor
NDUFB2	GOLPH3+3L Interactor
NDUFB4	GOLPH3+3L Interactor
NDUFB5	GOLPH3+3L Interactor
NDUFB6	GOLPH3+3L Interactor
NDUFB7	GOLPH3+3L Interactor
NDUFB9	GOLPH3+3L Interactor
NDUFC2	GOLPH3+3L Interactor
NDUFS1	GOLPH3+3L Interactor
NDUFS2	GOLPH3+3L Interactor
NDUFS3	GOLPH3+3L Interactor
NDUFS8	GOLPH3+3L Interactor
NDUFV2	GOLPH3+3L Interactor
NFS1	GOLPH3+3L Interactor
NGDN	GOLPH3+3L Interactor
NIPBL	GOLPH3+3L Interactor
NLRX1	GOLPH3+3L Interactor
NOC2L	GOLPH3+3L Interactor
NOC3L	GOLPH3+3L Interactor
NOC4L	GOLPH3+3L Interactor
NOP14	GOLPH3+3L Interactor
NOP9	GOLPH3+3L Interactor
NRM	GOLPH3+3L Interactor
NSUN5	GOLPH3+3L Interactor
NT5DC2	GOLPH3+3L Interactor
NUDC	GOLPH3+3L Interactor
NUDT21	GOLPH3+3L Interactor
NUP107	GOLPH3+3L Interactor
NUP188	GOLPH3+3L Interactor
NUP205	GOLPH3+3L Interactor
NUP93	GOLPH3+3L Interactor

OGFOD3	GOLPH3+3L Interactor
OPA1	GOLPH3+3L Interactor
OPA3	GOLPH3+3L Interactor
OSTC	GOLPH3+3L Interactor
OXA1L	GOLPH3+3L Interactor
P2RX4	GOLPH3+3L Interactor
PAM16	GOLPH3+3L Interactor
PARP16	GOLPH3+3L Interactor
PDF	GOLPH3+3L Interactor
PDS5A	GOLPH3+3L Interactor
PDS5B	GOLPH3+3L Interactor
PEX11G	GOLPH3+3L Interactor
PEX16	GOLPH3+3L Interactor
PEX3	GOLPH3+3L Interactor
PGS1	GOLPH3+3L Interactor
PHB	GOLPH3+3L Interactor
PHB2	GOLPH3+3L Interactor
PHGDH	GOLPH3+3L Interactor
PHKA1	GOLPH3+3L Interactor
PHKB	GOLPH3+3L Interactor
PHKG2	GOLPH3+3L Interactor
PIEZO1	GOLPH3+3L Interactor
PIGK	GOLPH3+3L Interactor
PIGS	GOLPH3+3L Interactor
PIGT	GOLPH3+3L Interactor
PIGU	GOLPH3+3L Interactor
PISD	GOLPH3+3L Interactor
PNPLA6	GOLPH3+3L Interactor
PNPLA8	GOLPH3+3L Interactor
POLR2H	GOLPH3+3L Interactor
PON2	GOLPH3+3L Interactor
PPM1L	GOLPH3+3L Interactor
PPP2R5C	GOLPH3+3L Interactor
PPP2R5D	GOLPH3+3L Interactor
PPP6R3	GOLPH3+3L Interactor
PRAF2	GOLPH3+3L Interactor
PREB	GOLPH3+3L Interactor
PTDSS2	GOLPH3+3L Interactor
PTGFRN	GOLPH3+3L Interactor
PTK7	GOLPH3+3L Interactor
PTPMT1	GOLPH3+3L Interactor
PTTG1IP	GOLPH3+3L Interactor
PXMP2	GOLPH3+3L Interactor
PXMP4	GOLPH3+3L Interactor
PXYLP1	GOLPH3+3L Interactor
PYCR1	GOLPH3+3L Interactor
RABL3	GOLPH3+3L Interactor
RALA;RALB	GOLPH3+3L Interactor
RBM26	GOLPH3+3L Interactor
RBM27	GOLPH3+3L Interactor
RBMX	GOLPH3+3L Interactor
RDH14	GOLPH3+3L Interactor
REEP2	GOLPH3+3L Interactor
REEP5	GOLPH3+3L Interactor
RFC3	GOLPH3+3L Interactor
RFT1	GOLPH3+3L Interactor
RHBDD1	GOLPH3+3L Interactor
RHOT1	GOLPH3+3L Interactor
RIF1	GOLPH3+3L Interactor
RNF126	GOLPH3+3L Interactor
RPL10	GOLPH3+3L Interactor
RPL10A	GOLPH3+3L Interactor
RPL23	GOLPH3+3L Interactor
RPL23A	GOLPH3+3L Interactor
RPL29	GOLPH3+3L Interactor
RPL32	GOLPH3+3L Interactor
RPL36	GOLPH3+3L Interactor
RPL38	GOLPH3+3L Interactor
RPLP2	GOLPH3+3L Interactor
RPS12	GOLPH3+3L Interactor
RPS16	GOLPH3+3L Interactor
RPS18	GOLPH3+3L Interactor
RPS19	GOLPH3+3L Interactor
RPS25	GOLPH3+3L Interactor
RPS27	GOLPH3+3L Interactor
RPS27A;UBB;UBC	GOLPH3+3L Interactor
RPS29	GOLPH3+3L Interactor
RPS3A	GOLPH3+3L Interactor
RPS5	GOLPH3+3L Interactor
RPSA	GOLPH3+3L Interactor
RQCD1	GOLPH3+3L Interactor
RRP12	GOLPH3+3L Interactor
RTCB	GOLPH3+3L Interactor
RUVBL1	GOLPH3+3L Interactor
SCAP	GOLPH3+3L Interactor
SCCPDH	GOLPH3+3L Interactor
SCD5	GOLPH3+3L Interactor
SCO1	GOLPH3+3L Interactor
SCO2	GOLPH3+3L Interactor
SEC61A1	GOLPH3+3L Interactor
SEC62	GOLPH3+3L Interactor
SEPN1	GOLPH3+3L Interactor
SERINC1	GOLPH3+3L Interactor
SETD9	GOLPH3+3L Interactor

SF3B1	GOLPH3+3L Interactor
SF3B2	GOLPH3+3L Interactor
SF3B6	GOLPH3+3L Interactor
SFT2D2	GOLPH3+3L Interactor
SFT2D3	GOLPH3+3L Interactor
SGPL1	GOLPH3+3L Interactor
SIGMAR1	GOLPH3+3L Interactor
SLC16A1	GOLPH3+3L Interactor
SLC19A1	GOLPH3+3L Interactor
SLC1A3	GOLPH3+3L Interactor
SLC1A5	GOLPH3+3L Interactor
SLC25A1	GOLPH3+3L Interactor
SLC25A11	GOLPH3+3L Interactor
SLC25A17	GOLPH3+3L Interactor
SLC25A19	GOLPH3+3L Interactor
SLC25A22;SLC25A18	GOLPH3+3L Interactor
SLC25A3	GOLPH3+3L Interactor
SLC25A4	GOLPH3+3L Interactor
SLC25A6	GOLPH3+3L Interactor
SLC27A3	GOLPH3+3L Interactor
SLC30A9	GOLPH3+3L Interactor
SLC35F2	GOLPH3+3L Interactor
SLC35F6	GOLPH3+3L Interactor
SLC3A2	GOLPH3+3L Interactor
SLC46A2	GOLPH3+3L Interactor
SLIRP	GOLPH3+3L Interactor
SLTM	GOLPH3+3L Interactor
SMIM1	GOLPH3+3L Interactor
SMIM20	GOLPH3+3L Interactor
SMPD4	GOLPH3+3L Interactor
SNRPC	GOLPH3+3L Interactor
SNRPD2	GOLPH3+3L Interactor
SNX14	GOLPH3+3L Interactor
SNX25	GOLPH3+3L Interactor
SORT1	GOLPH3+3L Interactor
SPPL2B	GOLPH3+3L Interactor
SPTLC2	GOLPH3+3L Interactor
SRP14	GOLPH3+3L Interactor
SRPRB	GOLPH3+3L Interactor
SSBP1	GOLPH3+3L Interactor
SSR1	GOLPH3+3L Interactor
SSR3	GOLPH3+3L Interactor
STEAP2	GOLPH3+3L Interactor
STEAP3	GOLPH3+3L Interactor
STK19	GOLPH3+3L Interactor
STOML2	GOLPH3+3L Interactor
STRBP	GOLPH3+3L Interactor
STT3A	GOLPH3+3L Interactor
STX12	GOLPH3+3L Interactor
STX18	GOLPH3+3L Interactor
STX4	GOLPH3+3L Interactor
STX7	GOLPH3+3L Interactor
STX8	GOLPH3+3L Interactor
STXBP3	GOLPH3+3L Interactor
SUN2	GOLPH3+3L Interactor
SYMPK	GOLPH3+3L Interactor
SYNE2	GOLPH3+3L Interactor
SYNE3	GOLPH3+3L Interactor
TAMM41	GOLPH3+3L Interactor
TARBP1	GOLPH3+3L Interactor
TBC1D10A	GOLPH3+3L Interactor
TBC1D10B	GOLPH3+3L Interactor
TBCD	GOLPH3+3L Interactor
TBL2	GOLPH3+3L Interactor
TBP;TBPL2	GOLPH3+3L Interactor
TBRG4	GOLPH3+3L Interactor
TCTA	GOLPH3+3L Interactor
TECR	GOLPH3+3L Interactor
TELO2	GOLPH3+3L Interactor
TENM3	GOLPH3+3L Interactor
TEX10	GOLPH3+3L Interactor
THEM6	GOLPH3+3L Interactor
TIMELESS	GOLPH3+3L Interactor
TIMM13	GOLPH3+3L Interactor
TIMM50	GOLPH3+3L Interactor
TIMM8A	GOLPH3+3L Interactor
TIMMDC1	GOLPH3+3L Interactor
TMCO1	GOLPH3+3L Interactor
TMEM106B	GOLPH3+3L Interactor
TMEM109	GOLPH3+3L Interactor
TMEM11	GOLPH3+3L Interactor
TMEM161A	GOLPH3+3L Interactor
TMEM192	GOLPH3+3L Interactor
TMEM201	GOLPH3+3L Interactor
TMEM206	GOLPH3+3L Interactor
TMEM245	GOLPH3+3L Interactor
TMEM246	GOLPH3+3L Interactor
TMEM35	GOLPH3+3L Interactor
TMEM41B	GOLPH3+3L Interactor
TMEM43	GOLPH3+3L Interactor
TMEM97	GOLPH3+3L Interactor
TMPO	GOLPH3+3L Interactor
TMX1	GOLPH3+3L Interactor
TNFRSF10B	GOLPH3+3L Interactor

TNFRSF10D	GOLPH3+3L Interactor
TNPO3	GOLPH3+3L Interactor
TOMM20	GOLPH3+3L Interactor
TOMM70A	GOLPH3+3L Interactor
TOR1AIP1	GOLPH3+3L Interactor
TP53I11	GOLPH3+3L Interactor
TPP1	GOLPH3+3L Interactor
TRABD	GOLPH3+3L Interactor
TRAM1	GOLPH3+3L Interactor
TRAPPC10	GOLPH3+3L Interactor
TRIP13	GOLPH3+3L Interactor
TSKU	GOLPH3+3L Interactor
TTC37	GOLPH3+3L Interactor
TTI1	GOLPH3+3L Interactor
TTI2	GOLPH3+3L Interactor
TUBA1A;TUBA3C	GOLPH3+3L Interactor
TUBA1B;TUBA4A	GOLPH3+3L Interactor
TUBB	GOLPH3+3L Interactor
TUBB2B	GOLPH3+3L Interactor
TUBB4A	GOLPH3+3L Interactor
TUBB4B	GOLPH3+3L Interactor
TUBB8	GOLPH3+3L Interactor
TUBG1;TUBG2	GOLPH3+3L Interactor
TVP23B;TVP23C	GOLPH3+3L Interactor
TYK2	GOLPH3+3L Interactor
U2AF1	GOLPH3+3L Interactor
UBE3C	GOLPH3+3L Interactor
UGGT2	GOLPH3+3L Interactor
UGT8	GOLPH3+3L Interactor
UNC45A	GOLPH3+3L Interactor
UNC50	GOLPH3+3L Interactor
UNC93B1	GOLPH3+3L Interactor
UQCC1	GOLPH3+3L Interactor
UQCC2	GOLPH3+3L Interactor
UQCR10	GOLPH3+3L Interactor
UQCRC1	GOLPH3+3L Interactor
UQCRC2	GOLPH3+3L Interactor
UQCRQ	GOLPH3+3L Interactor
USMG5	GOLPH3+3L Interactor
USP22	GOLPH3+3L Interactor
UTP20	GOLPH3+3L Interactor
VAMP2	GOLPH3+3L Interactor
VAMP3	GOLPH3+3L Interactor
VAMP4	GOLPH3+3L Interactor
VAMP8	GOLPH3+3L Interactor
VDAC1	GOLPH3+3L Interactor
VDAC2	GOLPH3+3L Interactor
VKORC1	GOLPH3+3L Interactor
VKORC1L1	GOLPH3+3L Interactor
VTI1A	GOLPH3+3L Interactor
VTI1B	GOLPH3+3L Interactor
VWA8	GOLPH3+3L Interactor
WWOX	GOLPH3+3L Interactor
XPO1	GOLPH3+3L Interactor
XPO5	GOLPH3+3L Interactor
XPO6	GOLPH3+3L Interactor
XPOT	GOLPH3+3L Interactor
XXYLT1	GOLPH3+3L Interactor
XYLT2	GOLPH3+3L Interactor
YBEY	GOLPH3+3L Interactor
YBX3	GOLPH3+3L Interactor
YWHAB	GOLPH3+3L Interactor
YWHAG	GOLPH3+3L Interactor
YWHAH	GOLPH3+3L Interactor
ZC3H4	GOLPH3+3L Interactor
ZNF787	GOLPH3+3L Interactor
ZW10	GOLPH3+3L Interactor
FAM114A2	FAM114A2 Interactor
EEF1B2	FAM114A2 Interactor
TMEM55B	FAM114A2 Interactor
CKAP4	FAM114A2 Interactor
ATP5D	FAM114A2 Interactor
GNB1	FAM114A2 Interactor
CARS	FAM114A2 Interactor
EEF1G	FAM114A2 Interactor
MGST3	FAM114A2 Interactor
LANCL2	FAM114A2 Interactor
HSPA8	FAM114A2 Interactor
TMEM143	FAM114A2 Interactor
C16orf13	FAM114A2 Interactor
HSPA9	FAM114A2 Interactor
CHD4	FAM114A2 Interactor
SLC27A1	FAM114A2 Interactor
NUMA1	FAM114A2 Interactor
RPS20	FAM114A2 Interactor
RPL18A	FAM114A2 Interactor
LMNB1	FAM114A2 Interactor
HSPA2	FAM114A2 Interactor
CBR3	FAM114A2 Interactor
TMEM55A	FAM114A2 Interactor
PLGRKT	FAM114A2 Interactor
PRPF3	FAM114A2 Interactor
EEF1A1P5;EEF1A1	FAM114A2 Interactor
SRPK1	FAM114A2 Interactor

PPIA	FAM114A2 Interactor
MRPL48	FAM114A2 Interactor
AAAS	COPI Proteome, GOLPH3+3L Interactor
APMAP	COPI Proteome, GOLPH3+3L Interactor
ARL6IP5	COPI Proteome, GOLPH3+3L Interactor
ATP1A1	COPI Proteome, GOLPH3+3L Interactor
ATP2A2	COPI Proteome, GOLPH3+3L Interactor
ATP2B1	COPI Proteome, GOLPH3+3L Interactor
B4GALT5	COPI Proteome, GOLPH3+3L Interactor
B4GAT1	COPI Proteome, GOLPH3+3L Interactor
BCAP31	COPI Proteome, GOLPH3+3L Interactor
C1GALT1	COPI Proteome, GOLPH3+3L Interactor
C3orf58	COPI Proteome, GOLPH3+3L Interactor
CALU	COPI Proteome, GOLPH3+3L Interactor
CANT1	COPI Proteome, GOLPH3+3L Interactor
CANX	COPI Proteome, GOLPH3+3L Interactor
CASC4	COPI Proteome, GOLPH3+3L Interactor
CHST14	COPI Proteome, GOLPH3+3L Interactor
CLPTM1L	COPI Proteome, GOLPH3+3L Interactor
CNIH4	COPI Proteome, GOLPH3+3L Interactor
CPD	COPI Proteome, GOLPH3+3L Interactor
CUX1	COPI Proteome, GOLPH3+3L Interactor
ECE1	COPI Proteome, GOLPH3+3L Interactor
EGFR	COPI Proteome, GOLPH3+3L Interactor
EMD	COPI Proteome, GOLPH3+3L Interactor
ESYT1	COPI Proteome, GOLPH3+3L Interactor
ESYT2	COPI Proteome, GOLPH3+3L Interactor
FAM3C	COPI Proteome, GOLPH3+3L Interactor
FUT8	COPI Proteome, GOLPH3+3L Interactor
GALNT1	COPI Proteome, GOLPH3+3L Interactor
GALNT10	COPI Proteome, GOLPH3+3L Interactor
GALNT7	COPI Proteome, GOLPH3+3L Interactor
GLCE	COPI Proteome, GOLPH3+3L Interactor
GLG1	COPI Proteome, GOLPH3+3L Interactor
GLT8D1	COPI Proteome, GOLPH3+3L Interactor
GOLGB1	COPI Proteome, GOLPH3+3L Interactor
GOLIM4	COPI Proteome, GOLPH3+3L Interactor
GOLT1B	COPI Proteome, GOLPH3+3L Interactor
GOSR1	COPI Proteome, GOLPH3+3L Interactor
HMOX1	COPI Proteome, GOLPH3+3L Interactor
IER3IP1	COPI Proteome, GOLPH3+3L Interactor
IMPAD1	COPI Proteome, GOLPH3+3L Interactor
KIAA0319L	COPI Proteome, GOLPH3+3L Interactor
KIAA2013	COPI Proteome, GOLPH3+3L Interactor
LDHB	COPI Proteome, GOLPH3+3L Interactor
LMAN2	COPI Proteome, GOLPH3+3L Interactor
LMBRD2	COPI Proteome, GOLPH3+3L Interactor
MAN1A2	COPI Proteome, GOLPH3+3L Interactor
MAN1B1	COPI Proteome, GOLPH3+3L Interactor
MAN2A1	COPI Proteome, GOLPH3+3L Interactor
MGAT4B	COPI Proteome, GOLPH3+3L Interactor
NCSTN	COPI Proteome, GOLPH3+3L Interactor
NPTN	COPI Proteome, GOLPH3+3L Interactor
P4HB	COPI Proteome, GOLPH3+3L Interactor
PGRMC1	COPI Proteome, GOLPH3+3L Interactor
PLOD2	COPI Proteome, GOLPH3+3L Interactor
QSOX2	COPI Proteome, GOLPH3+3L Interactor
RCN1	COPI Proteome, GOLPH3+3L Interactor
RER1	COPI Proteome, GOLPH3+3L Interactor
RPN1	COPI Proteome, GOLPH3+3L Interactor
RTN4	COPI Proteome, GOLPH3+3L Interactor
SCD	COPI Proteome, GOLPH3+3L Interactor
SCFD1	COPI Proteome, GOLPH3+3L Interactor
SEC22B	COPI Proteome, GOLPH3+3L Interactor
SERPINH1	COPI Proteome, GOLPH3+3L Interactor
SLC35E1	COPI Proteome, GOLPH3+3L Interactor
SLC39A14	COPI Proteome, GOLPH3+3L Interactor
SPTLC1	COPI Proteome, GOLPH3+3L Interactor
SSR4	COPI Proteome, GOLPH3+3L Interactor
ST3GAL1	COPI Proteome, GOLPH3+3L Interactor
ST3GAL4	COPI Proteome, GOLPH3+3L Interactor
STX10	COPI Proteome, GOLPH3+3L Interactor
SURF4	COPI Proteome, GOLPH3+3L Interactor
TGOLN2	COPI Proteome, GOLPH3+3L Interactor
TM9SF1	COPI Proteome, GOLPH3+3L Interactor
TM9SF3	COPI Proteome, GOLPH3+3L Interactor
TM9SF4	COPI Proteome, GOLPH3+3L Interactor
TMED1	COPI Proteome, GOLPH3+3L Interactor
TMED4	COPI Proteome, GOLPH3+3L Interactor
TMED9	COPI Proteome, GOLPH3+3L Interactor
TMEM167A	COPI Proteome, GOLPH3+3L Interactor
TMEM33	COPI Proteome, GOLPH3+3L Interactor
TPST1	COPI Proteome, GOLPH3+3L Interactor
UST	COPI Proteome, GOLPH3+3L Interactor
VAPA	COPI Proteome, GOLPH3+3L Interactor
YIF1B	COPI Proteome, GOLPH3+3L Interactor
YIPF3	COPI Proteome, GOLPH3+3L Interactor
ZFPL1	COPI Proteome, GOLPH3+3L Interactor
ABCD3	GOLPH3+3L Interactor, FAM114A2 Interactor
ATP5C1	GOLPH3+3L Interactor, FAM114A2 Interactor
COA3	GOLPH3+3L Interactor, FAM114A2 Interactor
COMTD1	GOLPH3+3L Interactor, FAM114A2 Interactor
CYC1	GOLPH3+3L Interactor, FAM114A2 Interactor
DDOST	GOLPH3+3L Interactor, FAM114A2 Interactor

DHCR24	GOLPH3+3L Interactor, FAM114A2 Interactor
ENPP1	GOLPH3+3L Interactor, FAM114A2 Interactor
EXOG	GOLPH3+3L Interactor, FAM114A2 Interactor
GALNT12	GOLPH3+3L Interactor, FAM114A2 Interactor
HIGD1A	GOLPH3+3L Interactor, FAM114A2 Interactor
IMMT	GOLPH3+3L Interactor, FAM114A2 Interactor
LEMD2	GOLPH3+3L Interactor, FAM114A2 Interactor
MOGS	GOLPH3+3L Interactor, FAM114A2 Interactor
OCIAD1	GOLPH3+3L Interactor, FAM114A2 Interactor
QPCTL	GOLPH3+3L Interactor, FAM114A2 Interactor
RPN2	GOLPH3+3L Interactor, FAM114A2 Interactor
SEC61B	GOLPH3+3L Interactor, FAM114A2 Interactor
SELH	GOLPH3+3L Interactor, FAM114A2 Interactor
SLC25A5	GOLPH3+3L Interactor, FAM114A2 Interactor
SLC27A4	GOLPH3+3L Interactor, FAM114A2 Interactor
TPST2	GOLPH3+3L Interactor, FAM114A2 Interactor
UXS1	GOLPH3+3L Interactor, FAM114A2 Interactor
YME1L1	GOLPH3+3L Interactor, FAM114A2 Interactor
ALDH3A2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
B3GAT3	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
CHPF2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
FAM20B	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
GALNT2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
GALNT3	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
GLB1L2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
GOLM1	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
HS2ST1	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
HSPA5	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
MANEAL	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
MGAT2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
POMGNT1	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
TMED10	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
TMED2	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor
YIPF6	COPI Proteome, GOLPH3+3L Interactor, FAM114A2 Interactor

COPI proteome from (Adolf et al., 2019), GOLPH3+3L interactors from Supplementary Table 8.5 and FAM114A2 interactors from Supplementary Table 8.2.

Supplementary Table 8.8 GST-CG9590 (1-332), CG9590-GST (1-332) vs GST

Protein name	Gene name	$-Log_{10}$ p-value	Difference Log_2 (CG9590/GST)
Q8I0G5	gammaCOP	6.458495735	11.6684138
Q7K7G0	deltaCOP	6.647906032	11.3263537
P45437	betaCop	4.234361666	9.01084963
Q9W0B8	alphaCOP	3.033614634	8.56966591
Q9VM94	homer	2.225568562	8.01177216
Q9W2S2;Q9VEG5	Atg8a;Atg8b	1.697159893	7.95775731
Q27601	CG2867;Prat	3.881162169	7.93734868
Q9VF39	CG9590	3.13154659	7.75931168
Q9Y0Y5	epsilonCOP	5.599719933	7.67093118
Q9W0M4	CG13887;Bap31	5.242821288	6.92852592
Q9VKU1	Fatp	4.797983267	6.78424199
Q9VFP0	CG3061	4.274719748	6.67615732
Q9VEX6	bor	4.528939095	6.61514982
C9QPI7	CG7461	5.993452116	6.46506055
Q7K110	OstDelta	5.816175795	6.42828433
Q24319	Ost48	5.327881865	6.34507052
Q9V455	Kap-alpha3	5.725648113	6.33814303
Q9Y140	CG7601	3.596159339	6.19795259
Q9W141	CG4692	5.486005876	6.03602378
Q9W0S7	Tudor-SN	2.629792862	5.94964027
Q9W2X6	ATPsyndelta	5.263040618	5.91531881
O18388	Fs(2)Ket	4.912006111	5.88667011
Q8SY53	Ostgamma	4.40889588	5.84554927
Q86NS4	Acsl	4.975486011	5.74611727
G4LU37	gammaCop-RA	2.8819094	5.73303795
Q7JZN0	Sec61beta	5.502277289	5.67894014
Q8SZ63	Golgin84	4.462303666	5.67594973
Q9VFE4	RpS5a	3.195079837	5.54158592
Q7K237	C1GalTA	6.135252709	5.50926844
O62621	betaCop	4.13559432	5.48984973
Q9W1G7	Nap1	2.992863228	5.44962215
O01666	ATPsyn-gamma	4.916264344	5.17445087
Q7KVX5	Vap-33A	4.565879672	5.08675607
Q27333	Alg3	5.646673235	5.04698245
Q76NQ0	CG33303	5.680505227	4.93471241
Q95U20	Opa1	3.711665426	4.89935112

Q9VG58	Hsp70	1.531124974	4.88502057
Q6WV17	pgant5	4.949586638	4.75428804
Q7KVA1	oxt	5.905071424	4.74656709
Q24439	Oscp	5.407479038	4.66454156
Q94516	ATPsyn-b	4.484835539	4.35097694
Q9XYN4	Ugt35a	3.612300087	4.29550775
Q9VRL0	Cyt-c1	5.500733649	4.26655579
Q9VRR4	Vap-33B	4.414068562	4.24488958
P91926	AP-2alpha	2.718594763	4.23985799
Q9VE98	CG8064	3.704280345	4.23424308
Q9VZF6	CG14997	4.491330196	4.17977269
Q24179	Slh	2.794046214	4.1417373
Q9Y117	pgant3	4.818961719	4.10987282
Q9VL69	CG5885	4.670644691	4.07337062
P20240	Ote	8.229058998	4.04769834
Q9VWG3	RpS10b	3.126486205	3.96120962
Q58SN2	Hmu	2.138514639	3.95185312
Q9VDS5	RhoGAP92B	1.828101259	3.81687609
Q9V3D9	Srp54k	3.447500697	3.81184038
Q9V9S8	FeCh	2.611089801	3.77680143
Q9VWL4	CG7556	3.782859065	3.74855391
Q7JW12	CG11007	2.892518145	3.73682976
Q86DS7	DnaJ-1	3.025797109	3.73133341
Q8STG9	Sec61alpha	2.349899337	3.69810073
Q24251	ATPsyn-d	2.292896602	3.68063577
O97067	CG1307	3.207482255	3.64896456
Q9VAY2	Gp93	1.686860932	3.62043985
Q9V3F2	lace	2.132536259	3.60221895
Q9VGH5	glo	1.534374346	3.508413
P55035	Rpn10	4.264379508	3.50270303
Q4V5K3	CG15653	3.426614786	3.50110118
Q9V4N3	dpld;Cyt-b5	4.525387153	3.48348173
Q4V607	RpS18	5.413634089	3.44127337
P22700	Ca-P60A	3.983890786	3.44018968
Q9VL22	CG4908	1.880911778	3.39252917
P52295	Pen	3.05742342	3.38997205
Q9U972	Desat1	1.631516511	3.33652719
Q9W3N1	spidey	3.722169954	3.27493445
Q9VLM5	l(2)k12914	4.0169813	3.26545461
P08928	Lam	1.615110077	3.20037874
Q9VV75	UQCR-C2	3.02470263	3.18182341
Q9VXQ5	Tcp-1zeta	4.495001543	3.18079917
Q9VSA3	CG12262	4.417291048	3.17590459
Q9VFV9	Droj2	4.633585035	3.15893491
Q9VZL3	Sc2;Sc2-RA	3.511848403	3.13451735
Q27237	l(2)tid;l(2)tid-RA;l(2)tid-RC	4.727116108	3.12591171
Q8SYL1	l(2)09851	4.171470906	3.09176668
Q9XZE4	Mpcp	3.081274649	3.06950982
Q9VBU6	RER1	4.164666053	3.01972103
Q8T6I0	Past1	4.470575715	2.97811572
Q24478	Cp190	2.365007024	2.87360096
Q9W414	Rpt4	1.608029686	2.82328033
Q2EJP2	Catsup	4.295929912	2.81839657
Q963E0	Cyp6a8	4.115460901	2.74999205
Q8IPG9	Bsg	1.756364685	2.74600824
P35128	ben	3.55557754	2.68805059
Q9V405	Rpt3	2.827423255	2.68357786
Q9VXT2	CG7872	2.733274763	2.65516376
Q9VHJ8	skap;CG11963	2.487533284	2.64845467
Q9VP27	Glg1	2.085064616	2.63407326
Q9VQX6	CG3702	3.194918803	2.624108
P41375	eIF-2beta	2.845857985	2.61494954
Q9VQX4	CG3714	1.959502032	2.59726397
Q9VHI1	hyx	2.482344621	2.57377243
Q9VF08	blp	2.330904376	2.5611817
Q9VV76	Syx8	1.705029778	2.55246099
Q9VPD3	ptdss1	4.444446293	2.53268115
Q9VYD5	CG1673	3.981952081	2.49422232
Q9I7K5	p24-2;eca	2.801840979	2.47672049
Q8IGH2	Rab11	1.833127964	2.47228622
Q8SXY6	bai	1.59971194	2.46495152

O77434	Sec22	1.815895551	2.43726953
Q9VYY2	CG1751;Spase25	1.869519357	2.43376446
Q7K0D8	Nup50	1.960449251	2.4206419
Q9VD26	CG5382	3.269969248	2.4137977
Q9VLC5	Aldh	1.935603108	2.39952151
Q9VUZ0	SsRbeta	6.313459416	2.38176632
Q8SYJ2	ND-MLRQ	3.123546693	2.3723882
P39018	RpS19a	2.344329141	2.34690444
O46084	Pgam5	1.592181968	2.27242756
Q53XD3	eIF-2gamma	2.607395251	2.23678144
Q9VZW7	CG2107	2.73136157	2.13087336
Q9VM95	Aatf	3.121683444	2.12194475
Q26365	sesB	2.575982645	2.09484069
Q9VVH5	UQCR-Q	3.542510561	2.0202926
P41374	eIF-2alpha	2.037871043	1.97135194
Q8MV48	GalNAc-T2	2.084910031	1.89640999
Q9VLC1	FucTB	3.210858613	1.83931414
P11147	Hsc70-4	3.237606415	1.82579867
P29844	Hsc70-3	3.150444229	1.82181676
Q7KLX3	Tapdelta	2.651571805	1.8074948
Q95SD0	Roe1	3.488428622	1.80581125
Q7K3G2	CG5742	2.094310905	1.76939265
Q9W5N2	RpL38	3.658130175	1.64120611
Q9NGK4	Mgat1	3.491143393	1.61037509
O62619	PyK	2.398832874	1.55944443
P0DKM0	Ccdc56	2.250597926	1.49083201
Q7K1V5	jagn	1.748663245	1.47553317
Q6W4K9	ATPsynbeta	4.140488887	1.41666444
Q7K161	IPP	2.136670462	1.37672551
A1ZAX1	eIF3-S8	2.202724428	1.37264188
Q8IHD2	RpS29	2.851993688	1.24688721
Q9VKX4	mRpS7	2.158399215	1.18709819
Q8MT18	CG5028	2.323898193	1.18592072
Q9VNA3	LD29996	2.880584812	1.1650308
P35381	blw	3.104406557	0.96560891
P55828	RpS20	2.284230372	0.89604568
P06605	alphaTub	2.572308007	0.58479945

Supplementary Table 8.9 Lectin affinity chromatography *ΔCG9590* vs wild-type adult flies

Protein ID	Gene Name	Jac Spectral Counts			VVA Spectral Counts			Relative Abundance Δ*CG9590*/WT	
		WT	#31	#76	WT	#31	#76	Jac	VVA
Q9W2L7_DROME	CG34396	14	4	7	8	3	1	0.393	0.250
A0A0B4K6Y7_DROME (+5)	pyd	23	10	8	13	3	3	0.391	0.231
E1JJ21_DROME	tmod	13	4	6	17	10	12	0.385	0.647
Q24160_DROME (+2)	Hmu	21	7	9	45	12	9	0.381	0.233
B3NVT4_DROER (+2)	AP-1-2beta	16	6	6	21	4	7	0.375	0.262
A0A0B4KGP8_DROME (+2)	eIF-2gamma	12	3	6	1	0	0	0.375	0.000
A0A0B4JCU7_DROME (+10)	Mmp1	11	5	3	9	3	1	0.364	0.222
Q7K0E6_DROME	Aats-asp	10	1	6	10	4	9	0.350	N/A
TSR1_DROME (+3)	CG7338	10	1	6	11	2	5	0.350	0.318
CEG1A_DROME (+3)	CenG1A	23	10	6	0	2	1	0.348	N/A
D1Z3A1_DROME (+6)	RpL28	9	1	5	7	4	8	0.333	0.857
Q7JXU4_DROME	Ns2	23	5	9	20	9	6	0.304	0.375
E1JJF4_DROME (+1)	sws	10	4	2	4	8	1	0.300	1.125
B7YZL1_DROME (+4)	CG10737	28	7	8	28	4	4	0.268	0.143
A1Z909_DROME	CG8545	42	5	16	57	18	34	0.250	0.456
A0T1Z4_DROME (+2)	yrt	15	5	2	13	1	0	0.233	0.038
B3N8T2_DROER (+4)	Rs1	13	3	3	10	3	4	0.231	0.350
RL4_DROME (+4)	RpL4	9	3	1	6	9	9	0.222	1.500
Q8SY39_DROME	CG6994	17	3	4	23	7	3	0.206	0.217
SSBP_DROME (+5)	mtSSB	10	1	3	10	1	0	0.200	0.050
Q9VFV1_DROME (+1)	yellow-e	13	2	3	20	4	5	0.192	0.225

M9MRJ4_DROME	Msp300	36	2	11	103	4	15	0.181	0.092
Q9VZV2_DROME (+5)	Cht7	39	10	4	16	4	1	0.179	0.156
Q9VBL3_DROME	CG5886	20	1	6	14	3	2	0.175	0.179
B4PR78_DROYA (+2)	Gp93	9	2	1	26	6	9	0.167	0.288
O76874_DROME (+1)	CG3777	34	7	4	27	15	11	0.162	0.481
Q9VFR3_DROME	Cht5	24	4	3	25	10	4	0.146	0.280
B3N7L4_DROER (+4)	CG8070	11	2	1	9	2	1	0.136	0.167
B7YZZ9_DROME (+1)	Rrp1	34	4	5	45	7	6	0.132	0.144
M9NCX5_DROME (+1)	Cda5	12	3	0	14	6	2	0.125	0.286
E4NKH2_DROME (+2)	CG7896	16	0	4	26	14	10	0.125	0.462
A0A0B4JD11_DROME (+1)	Ef1gamma	12	2	1	30	3	2	0.125	0.083
A0A0B4K7A5_DROME (+1)	Hts	113	12	16	139	21	24	0.124	0.162
Q0E9E2_DROME (+3)	PCB	50	4	8	20	14	10	0.120	0.600
A1Z9J3_DROME	shot	9	2	0	39	6	5	0.111	0.141
C5WLN0_DROME (+3)	CG2691	15	2	1	8	1	1	0.100	0.125
B4HQ54_DROSE (+2)	CG9143	21	0	4	6	0	4	0.095	0.333
LSP1G_DROME	Lsp1gamma	54	7	3	56	15	8	0.093	0.205
A0A0B4LHF7_DROME (+4)	pit	20	0	3	21	4	5	0.075	0.214
B5RJM2_DROME (+1)	CG4806	14	0	2	21	2	4	0.071	0.143
H0RNN8_DROME (+1)	CG13096	45	3	2	88	3	3	0.056	0.034
FBP1_DROME (+1)	Fbp1	10	1	0	15	4	0	0.050	0.133
B4IG55_DROSE (+3)	Nna1	15	1	0	2	0	0	0.033	0.000
LSP1A_DROME	Lsp1alpha	12	0	0	24	1	0	0.000	0.021
LSP1B_DROME (+1)	Lsp1beta	13	0	0	19	0	0	0.000	0.000
B4HU44_DROSE	Hexo1	14	0	0	7	0	0	0.000	0.000
A0A0B4LH29_DROME (+2)	CG44245	11	0	0	15	0	1	0.000	0.033
Q9VXY3_DROME	CG5599	39	0	0	9	0	0	0.000	0.000
A0A0B4KGX1_DROME (+2)	jar	10	0	0	23	9	9	0.000	0.391

Bibliography

Adolf, F., Rhiel, M., Hessling, B., Gao, Q., Hellwig, A., Bethune, J. and Wieland, F. T. (2019). Proteomic profiling of mammalian COPII and COPI vesicles. *Cell Rep* **26**, 250–265.

Ali, M. F., Chachadi, V. B., Petrosyan, A. and Cheng, P. W. (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.

Andag, U. and Schmitt, H. D. (2003). Dsl1p, an essential component of the Golgi-endoplasmic reticulum retrieval system in yeast, uses the same sequence motif to interact with different subunits of the COPI vesicle coat. *J Biol Chem* **278**, 51722–51734.

Antonny, B. (2011). Mechanisms of Membrane Curvature Sensing. *Annu. Rev. Biochem.* **80**, 101–123.

Appenzeller-Herzog, C. and Hauri, H. P. (2006). The ER-Golgi intermediate compartment (ERGIC): In search of its identity and function. *J. Cell Sci.* **119**, 2173–2183.

Baker, J. A., Wong, W. C., Eisenhaber, B., Warwicker, J. and Eisenhaber, F. (2017). Charged residues next to transmembrane regions revisited: “Positive-inside rule” is complemented by the “negative inside depletion/outside enrichment rule.” *BMC Biol.* **15**, 66.

Bard, F. and Chia, J. (2016). Cracking the Glycome Encoder: Signaling, Trafficking, and Glycosylation. *Trends Cell Biol* **26**, 379–388.

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Alpi, E., Bely, B., Bingley, M., Britto, R., Bursteinas, B., Busiello, G., et al. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515.

Baumann, N. A., Sullivan, D. P., Ohvo-Rekila, H., Simonot, C., Pottekat, A., Klaassen, Z., Beh, C. T. and Menon, A. K. (2005). Transport of newly synthesized sterol to the sterol-enriched plasma membrane occurs via nonvesicular equilibration. *Biochemistry* **44**, 5816–5826.

Beck, R., Sun, Z., Adolf, F., Rutz, C., Bassler, J., Wild, K., Sinning, I., Hurt, E., Brügger, B., Béthune, J., et al. (2008). Membrane curvature induced by Arf1-GTP is essential for vesicle formation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11731–11736.

Bigay, J., Casella, J. F., Drin, G., Mesmin, B. and Antonny, B. (2005). ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *EMBO J* **24**, 2244–2253.

Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H. and Hansson, G. C. (2015). New developments in goblet cell mucus secretion and function. *Mucosal Immunol.* **8**, 712–719.

Boncompain, G., Divoux, S., Gareil, N., de Forges, H., Lescure, A., Latreche, L., Mercanti, V., Jollivet, F., Raposo, G. and Perez, F. (2012). Synchronization of secretory protein traffic in populations of cells. *Nat Methods* **9**, 493–498.

Boulaflous, A., Saint-Jore-Dupas, C., Herranz-Gordo, M. C., Pagny-Salehabadi, S., Plasson, C., Garidou, F., Kiefer-Meyer, M. C., Ritzenthaler, C., Faye, L. and Gomord, V. (2009). Cytosolic N-terminal arginine-based signals together with a luminal signal target a type II membrane protein to the plant ER. *BMC Plant Biol* **9**, 144.

Bretscher, M. S. and Munro, S. (1993). Cholesterol and the Golgi apparatus. *Science* **261**, 1280–1281.

Brugger, B., Sandhoff, R., Wegehingel, S., Gorgas, K., Malsam, J., Helms, J. B., Lehmann, W. D., Nickel, W. and Wieland, F. T. (2000). Evidence for segregation of sphingomyelin and cholesterol during formation of COPI-coated vesicles. *J Cell Biol* **151**, 507–518.

Buchan, D. W. A. and Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.*

Burke, J., Pettitt, J. M., Schachter, H., Sarkar, M. and Gleeson, P. A. (1992). The transmembrane and flanking sequences of beta 1,2-N-acetylglucosaminyltransferase I specify medial-Golgi localization. *J Biol Chem* **267**, 24433–24440.

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

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

Campelo, F., van Galen, J., Turacchio, G., Parashuraman, S., Kozlov, M. M., Garcia-Parajo, M. F. and Malhotra, V. (2017). Sphingomyelin metabolism controls the shape and function of the Golgi cisternae. *Elife* **6**, e24603.

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

Cheong, F. Y., Sharma, V., Blagoveshchenskaya, A., Oorschot, V. M., Brankatschk, B., Klumperman, J., Freeze, H. H. and Mayinger, P. (2010). Spatial regulation of Golgi phosphatidylinositol-4-phosphate is required for enzyme localization and glycosylation fidelity. *Traffic* **11**, 1180–1190.

Chia, J., Tay, F. and Bard, F. (2019). The GalNAc-T activation (GAIA) pathway: Drivers and markers. *PLoS One* **14**, e0214118.

Ciechanover, A. and Ben-Saadon, R. (2004). N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol* **14**, 103–106.

Cosson, P., Lefkir, Y., Demolliere, C. and Letourneur, F. (1998). New COPI-binding motifs involved in ER retrieval. *EMBO J* **17**, 6863–6870.

Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805.

Diao, A., Rahman, D., Pappin, D. J., Lucocq, J. and Lowe, M. (2003). The coiled-coil membrane protein golgin-84 is a novel rab effector required for Golgi ribbon formation. *J Cell Biol* **160**, 201–212.

Dietrich, C., Bagatolli, L. A., Volovyk, Z. N., Thompson, N. L., Levi, M., Jacobson, K. and Gratton, E. (2001). Lipid rafts reconstituted in model membranes. *Biophys J* **80**, 1417–1428.

Dippold, H. C., Ng, M. M., Farber-Katz, S. E., Lee, S. K., Kerr, M. L., Peterman, M. C., Sim, R., Wiharto, P. A., Galbraith, K. A., Madhavarapu, S., et al. (2009). GOLPH3 bridges phosphatidylinositol-4-phosphate and actomyosin to stretch and shape the Golgi to promote budding. *Cell* **139**, 337–351.

Dodonova, S. O., Diestelkoetter-Bachert, P., von Appen, A., Hagen, W. J., Beck, R., Beck, M., Wieland, F. and Briggs, J. A. (2015). A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science* **349**, 195–198.

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

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.

Donohoe, B. S., Kang, B. H. and Staehelin, L. A. (2007). Identification and

characterization of COPIa- and COPIb-type vesicle classes associated with plant and algal Golgi. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 163–168.

Doucet, C. M., Esmery, N., De Saint-Jean, M. and Antonny, B. (2015). Membrane curvature sensing by amphipathic helices is modulated by the surrounding protein backbone. *PLoS One* **10**, e0137965.

Drin, G., Casella, J. F., Gautier, R., Boehmer, T., Schwartz, T. U. and Antonny, B. (2007). A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146.

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

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Ernst, A. M., Syed, S. A., Zaki, O., Bottanelli, F., Zheng, H., Hacke, M., Xi, Z., Rivera-Molina, F., Graham, M., Rebane, A. A., et al. (2018). S-Palmitoylation Sorts Membrane Cargo for Anterograde Transport in the Golgi. *Dev. Cell* **47**, 479–493.

Fabini, G., Freilinger, A., Altmann, F. and Wilson, I. B. H. (2001). Identification of core α1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from Drosophila melanogaster: Potential basis of the neural anti-horseradish peroxidase epitope. *J. Biol. Chem.* **276**, 28058–28067.

Ferez-Vilar, J. and Hill, R. L. (1999). The structure and assembly of secreted mucins. *J. Biol. Chem.* **274**, 31751–31754.

García-Nafría, J., Watson, J. F. and Greger, I. H. (2016). IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci. Rep.* **6**, 27459.

Gautier, R., Douguet, D., Antonny, B. and Drin, G. (2008). HELIQUEST: A web server to screen sequences with specific α-helical properties. *Bioinformatics* **24**, 2101–2102.

Gill, D. J., Chia, J., Senewiratne, J. and Bard, F. (2010). Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *J. Cell Biol.* **189**, 843–858.

Gillingham, A. K., Sinka, R., Torres, I. L., Lilley, K. S. and Munro, S. (2014). Toward a comprehensive map of the effectors of rab GTPases. *Dev Cell* **31**, 358–373.

Gillingham, A. K., Bertram, J., Begum, F. and Munro, S. (2019). In vivo identification of GTPase interactors by mitochondrial relocalization and proximity biotinylation. *Elife* **8**, e45916.

Glick, B. S. and Nakano, A. (2009). Membrane traffic within the Golgi apparatus. *Annu Rev Cell Dev Biol* **25**, 113–132.

Godl, K., Johansson, M. E. V., Lidell, M. E., Mörgelin, M., Karlsson, H., Olson, F. J., Gum, J. R., Kim, Y. S. and Hansson, G. C. (2002). The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J. Biol. Chem.* **277**, 47248–47256.

Gomez-Navarro, N. and Miller, E. (2016). Protein sorting at the ER-Golgi interface. *J. Cell Biol.* **216**, 769–778.

Graham, T. R. and Burd, C. G. (2011). Coordination of Golgi functions by phosphatidylinositol 4-kinases. *Trends Cell Biol* **21**, 113–121.

Hagen, K. G. T., Zhang, L., Tian, E. and Zhang, Y. (2009). Glycobiology on the fly: Developmental and mechanistic insights from Drosophila. *Glycobiology* **19**, 102–111.

Hartmann-Fatu, C., Trusch, F., Moll, C. N., Michin, I., Hassinen, A., Kellokumpu, S. and Bayer, P. (2015). Heterodimers of tyrosylprotein sulfotransferases suggest existence of a higher organization level of transferases in the membrane of the trans-golgi apparatus. *J. Mol. Biol.* **427**, 1404–1412.

Helenius, A. and Aebi, M. (2004). Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049.

Hennies, H. C., Kornak, U., Zhang, H., Egerer, J., Zhang, X., Seifert, W., Kühnisch, J., Budde, B., Nätebus, M., Brancati, F., et al. (2008). Gerodermia osteodysplastica is caused by mutations in SCYL1BP1, a Rab-6 interacting golgin. *Nat. Genet.* **40**, 1410–1412.

Herbomel, G. G., Rojas, R. E., Tran, D. T., Ajinkya, M., Beck, L. and Tabak, L. A. (2017). The GalNAc-T Activation Pathway (GALA) is not a general mechanism for regulating mucin-type O-glycosylation. *PLoS One* **12**, e0179241.
Higy, M., Junne, T. and Spiess, M. (2004). Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* **43**, 12716–12722.
Holthuis, J. C., Pomorski, T., Raggers, R. J., Sprong, H. and Van Meer, G. (2001). The organizing potential of sphingolipids in intracellular membrane transport. *Physiol Rev* **81**, 1689–1723.
Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832.
Isaji, T., Im, S., Gu, W., Wang, Y., Hang, Q., Lu, J., Fukuda, T., Hashii, N., Takakura, D., Kawasaki, N., et al. (2014). An oncogenic protein Golgi phosphoprotein 3 up-regulates cell migration via sialylation. *J Biol Chem* **289**, 20694–20705.
Jones, D. T. and Cozzetto, D. (2015). DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*.
Jungmann, J. and Munro, S. (1998). Multi-protein complexes in the cis Golgi of Saccharomyces cerevisiae with α-1,6-mannosyltransferase activity. *EMBO J.* **17**, 423–434.
Katoh, T. and Tiemeyer, M. (2013). The N's and O's of Drosophila glycoprotein glycobiology. *Glycoconj. J.* **30**, 57–66.
Keller, P., Semenza, G. and Shaltiel, S. (1995). Phosphorylation of the N-Terminal Intracellular Tail of Sucrase–Isomaltase by cAMP-Dependent Protein Kinase. *Eur. J. Biochem.* **233**, 963–968.
Kellokumpu, S., Hassinen, A. and Glumoff, T. (2016). Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. *Cell Mol Life Sci* **73**, 305–325.
Killian, J. A. (1998). Hydrophobic mismatch between proteins and lipids in membranes. *Bba-Rev Biomembr.* **1376**, 401–416.
Klemm, R. W., Ejsing, C. S., Surma, M. A., Kaiser, H. J., Gerl, M. J., Sampaio, J. L., de Robillard, Q., Ferguson, C., Proszynski, T. J., Shevchenko, A., et al. (2009). Segregation of sphingolipids and sterols during formation of secretory vesicles at the trans-Golgi network. *J Cell Biol* **185**, 601–612.
Kornfeld, R. and Kornfeld, S. (1985). Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664.
Kurokawa, K., Ishii, M., Suda, Y., Ichihara, A. and Nakano, A. (2013). Live cell visualization of Golgi membrane dynamics by super-resolution confocal live imaging microscopy. *Methods Cell Biol* **118**, 235–242.
Liu, L., Doray, B. and Kornfeld, S. (2018). Recycling of Golgi glycosyltransferases requires direct binding to coatomer. *Proc Natl Acad Sci U S A* **115**, 8984–8989.
Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L. and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 2193.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495.
Ma, W. and Goldberg, J. (2013). Rules for the recognition of dilysine retrieval motifs by coatomer. *EMBO J* **32**, 926–937.
Maccioni, H. J., Quiroga, R. and Ferrari, M. L. (2011). Cellular and molecular biology of glycosphingolipid glycosylation. *J Neurochem* **117**, 589–602.
Magdeleine, M., Gautier, R., Gounon, P., Barelli, H., Vanni, S. and Antonny, B. (2016). A filter at the entrance of the Golgi that selects vesicles according to size and bulk lipid composition. *Elife* **5**, e16988.
Malsam, J., Satoh, A., Pelletier, L. and Warren, G. (2005). Golgin Tethers Define Subpopulations of COPI Vesicles. **307**, 1095–1098.
McCormick, C., Duncan, G., Goutsos, K. T. and Tufaro, F. (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.

McMahon, H. T. and Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat Rev Mol Cell Biol* **12**, 517–533.

Meng, Y., Wang, L., Chen, D., Chang, Y., Zhang, M., Xu, J. J., Zhou, R. and Zhang, Q. Y. (2016). LAPTM4B: An oncogene in various solid tumors and its functions. *Oncogene* **35**, 6359–6365.

Michelsen, K., Yuan, H. and Schwappach, B. (2005). Hide and run. Arginine-based endoplasmic-reticulum-sorting motifs in the assembly of heteromultimeric membrane proteins. *EMBO Rep* **6**, 717–722.

Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G. and Engelman, D. M. (2004). Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc Natl Acad Sci U S A* **101**, 4083–4088.

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

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K. and Giraldez, A. J. (2015). CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988.

Munro, S. (1991). Sequences within and adjacent to the transmembrane segment of alpha-2,6-sialyltransferase specify Golgi retention. *EMBO J* **10**, 3577–3588.

Munro, S. (1995). An investigation of the role of transmembrane domains in Golgi protein retention. *EMBO J* **14**, 4695–4704.

Nakajima, M., Mizumoto, S., Miyake, N., Kogawa, R., Iida, A., Ito, H., Kitoh, H., Hirayama, A., Mitsubuchi, H., Miyazaki, O., et al. (2013). Mutations in B3GALT6, which encodes a glycosaminoglycan linker region enzyme, cause a spectrum of skeletal and connective tissue disorders. *Am J Hum Genet* **92**, 927–934.

Nguyen, A. T., Chia, J., Ros, M., Hui, K. M., Saltel, F. and Bard, F. (2017). Organelle specific O-glycosylation drives MMP14 activation, tumor growth, and metastasis. *Cancer Cell* **32**, 639–653.

Nikolovski, N., Rubtsov, D., Segura, M. P., Miles, G. P., Stevens, T. J., Dunkley, T. P., Munro, S., Lilley, K. S. and Dupree, P. (2012). Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiol* **160**, 1037–1051.

Nilsson, T., Jackson, M. and Peterson, P. A. (1989). Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic-reticulum. *Cell* **58**, 707–718.

Nilsson, T., Lucocq, J. M., Mackay, D. and Warren, G. (1991). The membrane spanning domain of beta-1,4-galactosyltransferase specifies trans Golgi localization. *EMBO J* **10**, 3567–3575.

Nilsson, T., Rabouille, C., Hui, N., Watson, R. and Warren, G. (1996). The role of the membrane-spanning domain and stalk region of N-acetylglucosaminyltransferase I in retention, kin recognition and structural maintenance of the Golgi apparatus in HeLa cells. *J Cell Sci* **109**, 1975–1989.

Orci, L., Schekman, R. and Perrelet, A. (1996). Interleaflet clear space is reduced in the membrane of COP I and COP II-coated buds/vesicles. *Proc Natl Acad Sci U S A* **93**, 8968–8970.

Osman, N., McKenzie, I. F., Mouhtouris, E. and Sandrin, M. S. (1996). Switching amino-terminal cytoplasmic domains of alpha(1,2)fucosyltransferase and alpha(1,3)galactosyltransferase alters the expression of H substance and Galalpha(1,3)Gal. *J Biol Chem* **271**, 33105–33109.

Owen, D. J. and Evans, P. R. (1998). A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* **282**, 1327–1332.

Pacheco, A. R., Lazarus, J. E., Sit, B., Schmieder, S., Lencer, W. I., Blondel, C. J., Doench, J. G., Davis, B. M. and Waldor, M. K. (2018). CRISPR screen reveals that

EHEC's T3SS and Shiga toxin rely on shared host factors for infection. *MBio* **9**, e01003-18.
Pantazopoulou, A. and Glick, B. S. (2019). A kinetic view of membrane traffic pathways can transcend the classical view of Golgi compartments. *Front. Cell Dev. Biol.* **7**, 153.
Parsons, H. T., Stevens, T. J., McFarlane, H. E., Vidal-Melgosa, S., Griss, J., Lawrence, N., Butler, R., Sousa, M. M. L., Salemi, M., Willats, W. G. T., et al. (2019). Separating Golgi Proteins from Cis to Trans Reveals Underlying Properties of Cisternal Localization. *Plant Cell* **31**, 2010–2034.
Patterson, G. H., Hirschberg, K., Polishchuk, R. S., Gerlich, D., Phair, R. D. and Lippincott-Schwartz, J. (2008). Transport through the Golgi apparatus by rapid partitioning within a two-phase membrane system. *Cell* **133**, 1055–1067.
Pereira, N. A., Pu, H. X., Goh, H. and Song, Z. (2014). Golgi phosphoprotein 3 mediates the Golgi localization and function of protein O-linked mannose beta-1,2-N-acetlyglucosaminyltransferase 1. *J Biol Chem* **289**, 14762–14770.
Pesch, Y. Y., Riedel, D., Patil, K. R., Loch, G. and Behr, M. (2016). Chitinases and Imaginal disc growth factors organize the extracellular matrix formation at barrier tissues in insects. *Sci. Rep.* **6**, 18340.
Phillips, R., Ursell, T., Wiggins, P. and Sens, P. (2009). Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**, 379–385.
Prinz, W. A. (2007). Non-vesicular sterol transport in cells. *Prog Lipid Res* **46**, 297–314.
Raghu, D., Mobley, R. J., Shendy, N. A. M., Perry, C. H. and Abell, A. N. (2019). GALNT3 Maintains the Epithelial State in Trophoblast Stem Cells. *Cell Rep* **26**, 3684–3697.
Rai, A., Goody, R. S. and Müller, M. P. (2019). Multivalency in Rab effector interactions. *Small GTPases* **10**, 40–46.
Ramsey, K. A., Rushton, Z. L. and Ehre, C. (2016). Mucin agarose gel electrophoresis: western blotting for high-molecular weight glycoproteins. *J. Vis. Exp.* **112**, e54153.
Rawet, M., Levi-Tal, S., Szafer-Glusman, E., Parnis, A. and Cassel, D. (2010). ArfGAP1 interacts with coat proteins through tryptophan-based motifs. *Biochem Biophys Res Commun* **394**, 553–557.
Reggiori, F. and Pelham, H. R. (2002). A transmembrane ubiquitin ligase required to sort membrane proteins into multivesicular bodies. *Nat Cell Biol* **4**, 117–123.
Rizzo, R., Russo, D., Kurokawa, K., Sahu, P., Lombardi, B., Supino, D., Zhukovsky, M., Vocat, A., Pothukuchi, P., Kunnathully, V., et al. (2019). The Glyco-enzyme adaptor GOLPH3 Links Intra-Golgi Transport Dynamics to Glycosylation Patterns and Cell Proliferation. *bioRxiv*.
Robinson, M. S. (2015). Forty Years of Clathrin-coated Vesicles. *Traffic* **16**, 1210–1238.
Roboti, P., Sato, K. and Lowe, M. (2015). The golgin GMAP-210 is required for efficient membrane trafficking in the early secretory pathway. *J Cell Sci* **128**, 1595–1606.
Rodríguez-Piñeiro, A. M., Van Der Post, S., Johansson, M. E. V., Thomsson, K. A., Nesvizhskii, A. I. and Hansson, G. C. (2012). Proteomic study of the mucin granulae in an intestinal goblet cell model. *J. Proteome Res.* **11**, 1879–1890.
Roux, A., Cuvelier, D., Nassoy, P., Prost, J., Bassereau, P. and Goud, B. (2005). Role of curvature and phase transition in lipid sorting and fission of membrane tubules. *EMBO J* **24**, 1537–1545.
Roux, K. J., Kim, D. I., Raida, M. and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801–810.
Rudd, P. M. and Dwek, R. A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol* **32**, 1–100.
Sato, T. (2014). *Lectin-probed western blot analysis*.
Sato, K., Sato, M. and Nakano, A. (2001). Rer1p, a retrieval receptor for endoplasmic reticulum membrane proteins, is dynamically localized to the Golgi apparatus by coatomer. *J Cell Biol* **152**, 935–944.
Schaecher, S. R., Diamond, M. S. and Pekosz, A. (2008). The Transmembrane Domain of the Severe Acute Respiratory Syndrome Coronavirus ORF7b Protein Is Necessary

and Sufficient for Its Retention in the Golgi Complex. *J. Virol.* **82**, 9477–9491.
Schmitz, K. R., Liu, J., Li, S., Setty, T. G., Wood, C. S., Burd, C. G. and Ferguson, K. M. (2008). Golgi localization of glycosyltransferases requires a Vps74p oligomer. *Dev Cell* **14**, 523–534.
Schneiter, R., Brugger, B., Sandhoff, R., Zellnig, G., Leber, A., Lampl, M., Athenstaedt, K., Hrastnik, C., Eder, S., Daum, G., et al. (1999). Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of the lipid molecular species composition of yeast subcellular membranes reveals acyl chain-based sorting/remodeling of distinct molecular species en route to the plasma membrane. *J. Cell Biol.* **146**, 741–754.
Schoberer, J., Liebminger, E., Vavra, U., Veit, C., Grunwald-Gruber, C., Altmann, F., Botchway, S. W. and Strasser, R. (2019). The Golgi localization of GnTI requires a polar amino acid residue within Its transmembrane domain. *Plant Physiol* **180**, 859–873.
Schwientek, T., Mandel, U., Roth, U., Müller, S. and Hanisch, F. G. (2007). A serial lectin approach to the mucin-type O-glycoproteome of Drosophila melanogaster S2 cells. *Proteomics* **7**, 3264–3277.
Scott, K. L., Kabbarah, O., Liang, M. C., Ivanova, E., Anagnostou, V., Wu, J., Dhakal, S., Wu, M., Chen, S., Feinberg, T., et al. (2009). GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. *Nature* **459**, 1085–1090.
Sharpe, H. J., Stevens, T. J. and Munro, S. (2010). A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169.
Shikano, S. and Li, M. (2003). Membrane receptor trafficking: evidence of proximal and distal zones conferred by two independent endoplasmic reticulum localization signals. *Proc Natl Acad Sci U S A* **100**, 5783–5788.
Shin, J. J. H., Gillingham, A. K., Begum, F., Chadwick, J. and Munro, S. (2017). TBC1D23 is a bridging factor for endosomal vesicle capture by golgins at the trans-Golgi. *Nat. Cell Biol.* **19**, 1424–1432.
Sievers, F. and Higgins, D. G. (2014). Clustal Omega. *Curr. Protoc. Bioinforma.* **48**, 1.25.1- 1.25.33.
Sönnichsen, B., Lowe, M., Levine, T., Jämsä, E., Dirac-Svejstrup, B. and Warren, G. (1998). A role for giantin in docking COPI vesicles to Golgi membranes. *J. Cell Biol.* **140**, 1013–1021.
Sowa, M. E., Bennett, E. J., Gygi, S. P. and Harper, J. W. (2009a). Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **138**, 389–403.
Sowa, M. E., Bennett, E. J., Gygi, S. P. and Harper, J. W. (2009b). c. *Cell* **138**, 389–403.
Suckling, R. J., Poon, P. P., Travis, S. M., Majoul, I. V, Hughson, F. M., Evans, P. R., Duden, R. and Owen, D. J. (2015). Structural basis for the binding of tryptophan-based motifs by delta-COP. *Proc Natl Acad Sci U S A* **112**, 14242–14247.
Swift, A. M. and Machamer, C. E. (1991). A Golgi retention signal in a membrane-spanning domain of coronavirus E1 protein. *J Cell Biol* **115**, 19–30.
Tanaka, A., Tumkosit, U., Nakamura, S., Motooka, D., Kishishita, N., Priengprom, T., Sa-Ngasang, A., Kinoshita, T., Takeda, N. and Maeda, Y. (2017). Genome-wide screening uncovers the significance of N-sulfation of heparan sulfate as a host cell factor for chikungunya virus infection. *J Virol* **91**, e00432-17.
Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., et al. (2017). A subcellular map of the human proteome. *Science.* **356**, eaal3321.
Tian, S., Muneeruddin, K., Choi, M. Y., Tao, L., Bhuiyan, R. H., Ohmi, Y., Furukawa, K., Furukawa, K., Boland, S., Shaffer, S. A., et al. (2018). Genome-wide CRISPR screens for Shiga toxins and ricin reveal Golgi proteins critical for glycosylation. *PLoS Biol* **16**, e2006951.
Tie, H. C., Ludwig, A., Sandin, S. and Lu, L. (2018). The spatial separation of processing and transport functions to the interior and periphery of the Golgi stack. *Elife* **7**, e41301.
Tong, Z., Kim, M. S., Pandey, A. and Espenshade, P. J. (2014). Identification of

candidate substrates for the Golgi Tul1 E3 ligase using quantitative diGly proteomics in yeast. *Mol Cell Proteomics* **13**, 2871–2882.
Tu, L. and Banfield, D. K. (2010). Localization of Golgi-resident glycosyltransferases. *Cell Mol Life Sci* **67**, 29–41.
Tu, L., Tai, W. C., Chen, L. and Banfield, D. K. (2008). Signal-mediated dynamic retention of glycosyltransferases in the Golgi. *Science* **321**, 404–407.
Tu, L., Chen, L. and Banfield, D. K. (2012). A conserved N-terminal arginine-motif in GOLPH3-family proteins mediates binding to coatomer. *Traffic* **13**, 1496–1507.
Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M. and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.
Uemura, T. and Waguri, S. (2020). Emerging roles of Golgi/endosome-localizing monomeric clathrin adaptors GGAs. *Anat. Sci. Int.* **95**, 12–21.
Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science.* **347**, 1260419.
van Galen, J., Campelo, F., Martinez-Alonso, E., Scarpa, M., Martinez-Menarguez, J. A. and Malhotra, V. (2014). Sphingomyelin homeostasis is required to form functional enzymatic domains at the trans-Golgi network. *J Cell Biol* **206**, 609–618.
van Meel, E., Qian, Y. and Kornfeld, S. A. (2014). Mislocalization of phosphotransferase as a cause of mucolipidosis III αβ. *Proc Natl Acad Sci U S A* **111**, 3532–3537.
van Meer, G. (1989). Lipid traffic in animal cells. *Annu Rev Cell Biol* **5**, 247–275.
van Meer, G. (1998). Lipids of the Golgi membrane. *Trends Cell Biol* **8**, 29–33.
van Meer, G., Voelker, D. R. and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112–124.
von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **5**, 3021–3027.
Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., et al. (1997). Substrate specificities of three members of the human UDP-N-acetyl-α- D-galactosamine:polypeptide N-acetylgalactosaminyltransferase family, GalNAc- T1, -T2, and -T3. *J. Biol. Chem.* **272**, 23503–23514.
Wang, P., Wang, H., Gai, J., Tian, X., Zhang, X., Lv, Y. and Jian, Y. (2017). Evolution of protein N-glycosylation process in Golgi apparatus which shapes diversity of protein N-glycan structures in plants, animals and fungi. *Sci Rep* **7**, 40301.
Wassie, A. T., Zhao, Y. and Boyden, E. S. (2019). Expansion microscopy: principles and uses in biological research. *Nat Methods* **16**, 33–41.
Welch, L. G. and Munro, S. (2019). A tale of short tails, through thick and thin: investigating the sorting mechanisms of Golgi enzymes. *FEBS Lett.* **593**, 2452–2465.
Wilson, I. B. H. (2002). Functional characterization of Drosophila melanogaster peptide O-xylosyltransferase, the key enzyme for proteoglycan chain initiation and member of the core 2/I N-acetylglucosaminyltransferase family. *J. Biol. Chem.* **277**, 21207–21212.
Witkos, T. M., Chan, W. L., Joensuu, M., Rhiel, M., Pallister, E., Thomas-Oates, J., Mould, A. P., Mironov, A. A., Biot, C., Guerardel, Y., et al. (2019). GORAB scaffolds COPI at the trans-Golgi for efficient enzyme recycling and correct protein glycosylation. *Nat. Commun.* **10**, 127.
Wong, M. and Munro, S. (2014). Membrane trafficking. The specificity of vesicle traffic to the Golgi is encoded in the golgin coiled-coil proteins. *Science* **346**, 1256898.
Wong, M., Gillingham, A. K. and Munro, S. (2017). The golgin coiled-coil proteins capture different types of transport carriers via distinct N-terminal motifs. *BMC Biol* **15**, 3.
Woo, C. H., Gao, C., Yu, P., Tu, L., Meng, Z., Banfield, D. K., Yao, X. and Jiang, L. (2015). Conserved function of the lysine-based KXD/E motif in Golgi retention for endomembrane proteins among different organisms. *Mol Biol Cell* **26**, 4280–4293.

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

Wood, C. S., Hung, C. S., Huoh, Y. S., Mousley, C. J., Stefan, C. J., Bankaitis, V., Ferguson, K. M. and Burd, C. G. (2012). Local control of phosphatidylinositol 4-phosphate signaling in the Golgi apparatus by Vps74 and Sac1 phosphoinositide phosphatase. *Mol Biol Cell* **23**, 2527–2536.

Yamaji, T., Sekizuka, T., Tachida, Y., Sakuma, C., Morimoto, K., Kuroda, M. and Hanada, K. (2019). A CRISPR screen identifies LAPTM4A and TM9SF proteins as glycolipid-regulating factors. *Iscience* **11**, 409–424.

Yamamoto-Hino, M., Abe, M., Shibano, T., Setoguchi, Y., Awano, W., Ueda, R., Okano, H. and Goto, S. (2012). Cisterna-specific localization of glycosylation-related proteins to the Golgi apparatus. *Cell Struct. Funct.* **37**, 55–63.

Zhang, L. and Ten Hagen, K. G. (2019). O-Linked glycosylation in Drosophila melanogaster. *Curr. Opin. Struct. Biol.* **37**, 1–14.