TITLE

Separating Golgi proteins from cis to trans reveals underlying properties of cisternal localization

SHORT TITLE

Separating Golgi proteins from cis to trans

AUTHOR LIST

Harriet T. Parsons[1,4†*], Tim J. Stevens[2†], Heather E. McFarlane[3], Silvia Vidal-Melgosa[4], Johannes Griss[8,9], Nicola Lawrence[10], Richard Butler[10], Mirta M. L. Sousa[5], Michelle Salemi[7], William G. T. Willats[4], Christopher J. Petzold[6], Joshua L. Heazlewood[3], Kathryn S. Lilley[1]

1. Department of Biochemistry, Cambridge University, Cambridge, UK
2. MRC Laboratory of Molecular Biology, Cambridge, UK
3. School of Biosciences, University of Melbourne, Melbourne, Australia
4. Department of Plant and Environmental Sciences, Copenhagen University, Copenhagen, Denmark
5. Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway.
6. Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, USA
7. Proteomics Core Facility, University of California, Davis, USA
8. Division of Immunology, Allergy and Infectious Diseases, Department of Dermatology, Medical University of Vienna, Austria.
9. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK.
10. The Wellcome Trust and Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge 2 1QN, UK

[†]These authors contributed equally

*Correspondence: tempeparsons@gmail.com

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Harriet T. Parsons (tempeparsons@gmail.com).

ABSTRACT

The order of enzymatic activity across Golgi cisternae is essential for complex molecule biosynthesis. However, an inability to separate Golgi cisternae has meant the cisternal distribution of most resident proteins, and their underlying localization mechanisms, are unknown. Here, we exploit differences in surface charge of intact cisternae to perform the first separation of early to late Golgi sub-compartments. We determine protein and glycan abundance profiles across the Golgi; over 390 resident proteins are identified, including 136 new additions, with over 180 cisternal assignments. These assignments provide a means to better understand the functional roles of Golgi proteins and how they operate sequentially. Protein and glycan distributions are validated in-vivo, using high resolution microscopy. Results reveal distinct functional compartmentalization among resident Golgi proteins. Analysis of transmembrane proteins shows several sequence-based characteristics relating to pI, hydrophobicity, Ser abundance and Phe bilayer asymmetry that change across the Golgi. Overall this suggests a continuum of TM features, rather than discrete rules, which guide proteins to earlier or later locations within the Golgi stack.

INTRODUCTION

The Golgi is an ancient organelle, common to all eukaryotic lineages (1), consisting of a stack of flattened, membranous discs, or cisternae, in which protein and lipid cargoes are modified in a progressive manner, and substituted with complex glycan side chains (2–4). The Golgi is the hub of the secretory pathway, trafficking cargo-containing vesicles to and from the endoplasmic reticulum (ER) at the cis face (5) and to other cellular destinations at the trans face (6). There have been important advances in understanding trafficking processes from the trans-Golgi network (TGN) to post-TGN destinations (7–9), and many regulatory components of ER to cis-Golgi traffic have been determined (5,10). However, our understanding of the trafficking pathways within

the Golgi stack itself, and the mechanisms underlying spatial partitioning of proteins within stacks, is still somewhat limited.

Studying secretory organelle organization not only contributes to a general understanding of biochemical pathways and how protein localization is specified but also gives us the capacity to better control the complex, sequential biochemistry and trafficking processes of cellular secretion. Although understanding of how sequence characteristics localize proteins to organelles has advanced (11), no general sequence-based determinants of Golgi cisternal membrane localization are known (12). TM span length, retrieval and retention motifs (13–16) cannot sufficiently explain the distribution of resident proteins within the Golgi, implicating undiscovered factors governing intra-Golgi protein localization. Cutting-edge microscopy has localized a limited number of Golgi proteins (17), though tagging membrane proteins can increase aberrant localization (18). Consequently, too few proteins have been accurately localized within the Golgi to identify cisternal targeting sequences or map intra-Golgi trafficking pathways.

Modern mass spectrometry, using multiple separation stages and peptide mass fingerprinting, provides a way of simultaneously detecting and quantifying the occurrence of thousands of proteins in purified and enriched samples. This has allowed the compilation of proteome sets for sub-cellular compartments. Generally, these comparative proteomic analyses, which have proved essential to our understanding of vesicular trafficking (9,19), depend on some degree of physical separation of compartments. Here, the LOPIT (localization of organelle proteins by isotope tagging) technique, using density gradient centrifugation, has become the gold-standard for sub-cellular proteome discovery (20) and has provided ER, Golgi and TGN proteomes in Arabidopsis (21–23). However, to-date, only electrophoresis techniques have delivered adequate separation of Golgi cisternae. Free-flow electrophoresis (FFE) has been shown to separate vesicles according to small differences in surface charge (24,25). Although early attempts to separate the ER, Golgi cisternae and TGN using FFE were promising (26), contemporary technical limitations prevented proper follow-up and validation. In this study, we separate the Golgi sub-compartments in an

endomembrane-enriched sample from an Arabidopsis cell-suspension culture using FFE.

Plant suspension-culture cells are an attractive option for studying the endomembrane as they generate large quantities of intact Golgi cisternae (27). Centrifugation and gentle manipulation under negative pressure efficiently unstacks cisternae, which can be enriched on a simple step-gradient. A gradient of surface charge, likely resulting from flipping of negatively charged phospholipids to the outer leaflet, exists between the ER, Golgi, TGN and PM (26,27), and appears to exist across Golgi cisternae, which facilitates electrophoretic separation.

Here we combine gentle electrophoretic fractionation of largely intact endomembranes with high-throughput mass spectrometry, bioinformatics and imaging techniques to create one of the largest experimental data sets in this field to-date. We use both LOPIT and FFE abundance profiles to determine the localization of hundreds of resident proteins, protein cargo and glycan cargo through the secretory pathway at sub-Golgi resolution. Our approach is validated in serval ways, including using glycan immuno-gold transmission electron microscopy and protein fluorescence microscopy. We show sub-Golgi categorisations that are consistent with the progressive glycosylation functions of the Golgi. This then allows us to bioinformatically analyse sub-Golgi specific protein sequences to discover any trends or rules which may contribute to cisternal localization.

RESULTS

Experimental inputs

Using free-flow electrophoresis (FFE) we separated an endomembrane-enriched homogenate into 96 fractions according to surface charge. For each replicate sample, approximately 45 fractions with significant endomembrane protein content were selected in each case and analysed using shotgun proteomic mass spectrometry, to gauge the identity and relative amount of each protein in each fraction. A schematic

representation of our approach, using gentle separation of intact-membrane samples, mass-spectrometric proteomic identification and subsequent abundance profile generation is illustrated in Figure 1.

Preliminary investigations with two biological replicate samples (R1 and R2), performed with an ABSciex 5600 TripleTOF, identified over 1500 proteins and established the basic utility of our approach (and R1, which contained more material than R2, was later used for glycan/carbohydrate analysis). This was then followed-up with three high-sensitivity replicates (R3, R4 and R5), using an Orbitrap QExactive mass spectrometer, which detected over 2700 proteins and formed the basis of our main analysis.

Establishing updated sub-proteomes for the Golgi and other organelles

Before we could begin to dissect any cisternal separation of Golgi proteins, our first task was to establish updated protein sets of resident proteins for the Golgi and other membrane-bound compartments within our cell line. Current plant protein annotations sometimes contain contradictory locational information, often with no indication of which proteins are organelle residents or localize to multiple organelles. This is problematic when analyzing the Golgi, as distinguishing between cisternal residents, cargo, and vesicular proteins, is essential. It was especially important to generate accurate, updated ER and TGN proteomes; the ER showed the closest degree of FFE fraction overlap with the Golgi (see Figure 2a) and dual-localized ER-Golgi proteins were expected. Electrophoretic migration of the TGN was difficult to distinguish from the Golgi, as TGN proteins are both trafficked through and exchanged with the Golgi. Hence, updating the TGN proteome enabled TGN cargo to be distinguished from Golgi residents.

To date, the only proteomics technique capable of distinguishing resident and cargo proteins is LOPIT. In LOPIT, organelles are separated on a linear density gradient, fractions of which are labelled using isobaric tags. Tagging enables very accurate quantitation of protein abundances along the gradients. Proteins from the same organelle have similar abundance profiles so when, for example, principal component analysis is applied to quantitation data, organelle residents form distinct clusters and

multi-localized proteins do not. LOPIT was originally developed and validated using Arabidopsis, over a decade ago (22). Thorough cross-validation using immunoblots and imaging, as well as technical and bioinformatic updates, have led to LOPIT becoming the technique of choice for high-accuracy, whole-cell proteomics analysis (20,28–31) but it has never been re-applied to a whole-cell analysis of Arabidopsis. Updating resident organelle proteomes was therefore an essential first step in this study.

Multi-class support vector machine (SVM)-based methods are frequently combined with LOPIT to classify proteins according to their location (20,32,33). Here, we used proteins with clearly annotated localisation derived from the subcellular localization database for Arabidopsis proteins (SUBA) (34) and from (21) as the initial classification inputs (Table S1). This created organelle-specific clusters by partitioning the LOPIT profile data (i.e. density centrifugation profiles) according to the consensus of the initial markers. Classification parameters (see experimental procedures) were set such that organelle clusters remained tight and were therefore most likely to contain only resident proteins. When compared against fluorescent protein localization records housed in SUBA, <5% of proteins showed conflicting localizations. Given that the Golgi has been subject to relatively few proteomic studies, it was desirable to increase the number of known Golgi resident proteins. Hence, the SVM classification parameters were relaxed to permit <2% conflicts. This did not affect the tightness of the Golgi cluster, meaning accuracy was not compromised. For all organelles, proteins were only selected if present in 2 or more replicates.

Principal component analysis (PCA) revealed tight, distinct clusters for all subcellular compartments (Figure 2a). The compartments could be largely, but not entirely, separated by projection on to only two principle components. Hence, results were also visualized using t-distributed stochastic neighbour embedding (t-SNE), which attempts to combine data from all dimensions to a two-dimensional plot (35). t-SNE confirmed that clusters overlapping in Figure 2a, including the ER, Golgi and TGN, were indeed separate (Figure 2b). Importantly, for our later analyses, the TGN group was entirely distinct from the Golgi.

LOPIT resulted in the identification of 345 ER, 46 TGN and 397 Golgi resident proteins in three spatially distinct clusters, along with comprehensive lists of resident protein markers for all other organelles (Figure S1, Table S1, Supplemental Results). The currently annotated Arabidopsis Golgi proteome (covering all cell types) is estimated at ~ 530 proteins (36), suggesting that we identified a large majority of resident Golgi proteins present in our cell line.

Organelle FFE protein abundance profiles

Having established updated, resident proteomes for all major subcellular compartments we then used this to analyse FFE data. After merging high-sensitivity proteomic data from replicates R3-5 (see Methods) the combined, average FFE profiles of proteins previously known to reside in the ER and Golgi are illustrated in Figure 2c, alongside profiles for the newly-assigned ER and Golgi sets from LOPIT; the newly assigned proteins had remarkably similar profiles to those of established residents. Additionally, the combined FFE profiles for all other LOPIT sub-compartment classifications (Figure 2c) show that this data can be used to categorize non-endomembrane proteins as either non-secretory contaminants or cargo. Contaminants e.g. peroxisome, plus most chloroplast and PM proteins, had electrophoretic profiles similar to those observed in previous electrophoresis separations (37,38). Interestingly, some chloroplast, PM, vacuole and mitochondrial proteins had flat profiles which did not correspond to previous observations for those organelles (25,37). The subpopulation of proteins from these organelles with flat profiles were disproportionately enriched in features consistent with cargo subjected to post-translational modifications in the Golgi. Over 40% of non-Golgi proteins identified in Golgi-enriched fractions had either been found previously in vesicular trafficking proteomes (9), were S-acylated (9), contained a high-confidence N-glycosylation site (39) or had an experimentally determined glycosylphosphatidylinositol (GPI)-anchor (40).

Overall, Golgi proteins were detected across the entire region of the selected membrane fractions and did not obviously separate into discrete surface-charge regions (e.g. corresponding to different cisternae). However, even with separate sub-Golgi proteomes we would expect a somewhat overlapped situation here given that resident

proteins transit though, and possibly recycle, via adjacent compartments. Additionally, we are studying a superposition of different cellular and vesicular states, i.e. with varying surface charge.

As illustrated in Figure 2d for high-sensitivity replicates R3-R5, protein profiles were hierarchically clustered according to the pattern of their merged FFE abundance along the separated fractions. This clustering effectively pairs the most similar abundance profiles, in a progressive manner, and allowed us to visualize any innate groups that may occur within the FFE data, i.e. which may correspond to different organelles and sub-compartments. Given that Golgi cisternae remained largely intact during the FFE separation we did not directly separate Golgi residents from trafficking cargo, even if we might expect resident and cargo proteins to have different, characteristic FFE profiles. Hence, to objectively assign organelle residents to FFE profile clusters with highest confidence, we used only resident proteins from Figure 2a, b and Table S1.

When proteins with existing organelle annotations are compared by hierarchical clustering (Figure 2d and S2b) the grouped profiles clearly correspond to three major clusters: Golgi/TGN, ER and plasma membrane (PM), which have peak abundances in different regions of the FFE profile. Overall, the Golgi/TGN proteins tend to peak in early fractions (nearer anode), ER residents come in the middle and plasma membrane proteins come later. These features were also obvious in hierarchical clustering of the individual FFE replicate datasets R3-R5 (Figure S3), although they are clearest in the combined data, as expected. Also, looking within the large Golgi/TGN cluster we can see that TGN annotations largely group together and Golgi sub-clusters are present. Although the TGN FFE profiles are similar to, and hence cluster with, those from the Golgi, this presents no problems for our analysis as these compartments are entirely separate in the density centrifugation (LOPIT) analysis.

Dissecting the clustering further into minor sub-clusters which we label ***A***-***H***, we can see that ER proteins were distributed over a larger cluster ***E*** and a smaller, higher-variance cluster ***F*** (Figure 2d). Cluster ***D*** contained ER and Golgi proteins with profiles intermediate to most ER and Golgi proteins, possibly indicating a dual-localized group. Golgi proteins could be grouped into three mains clusters which appeared to form a

continuum along the electrophoretic gradient. The Golgi cluster with peak abundance closest to the anode (cluster ***A***) exhibited a zone of main protein abundance that was focused over a smaller number of fractions compared to e.g. cluster ***C***, which was wider and peaked closer to the cathode. Clusters ***G*** and ***H*** comprised mainly PM proteins and migrated furthest towards the cathode. This is consistent with previous reports that PM vesicles come out further towards that cathode than other endomembrane compartments (38).

Evidence for sub-Golgi separation in FFE profiles

To investigate whether the Golgi sub-clusters found in the FFE profiles had any correspondence with Golgi cisternae we performed an analysis of glycans in the FFE fractions that was coupled to electron microscopy of individual cisternae and also looked at proteins with well-established cisternal identity.

Cisternal polysaccharide distribution

Using immuno-gold transmission electron microcopy (TEM), we performed an in-situ analysis of glycan epitopes. These epitopes represented polysaccharides with different structural complexities, as would be found across the range of Golgi cisternae (see Table S2 for details). By using TEM on samples with gold-labelled antibodies we localized the glycans to individual cisternae with high spatial resolution.

As expected from previous analyses, glycan epitopes showed specific localizations for different Golgi membranes, with more structurally complex polysaccharides being associated with later cisternae (Figure 3a and summarized in Table S2). The overall TEM results (Figures 3a and 3b) are summarized as follows: 1) Anti-extensin LM1 was detected in the cis-Golgi; extensins have protein backbones, which provide a substrate for modification immediately after entering the Golgi. 2) Anti-mannan antibody LM21 was detected over cis and medial cisternae. 3) Antibodies for LM19, which recognizes partially methyl-esterified homogalacturonan (HG), and LM15, which recognizes a simply-branched, xylose-substituted epitope of xyloglucan (XG) , occur early but overall have a medial distribution and peak before XG epitopes with longer side-chains. 4) Anti-xyloglucan M87, which recognizes XG epitopes with medium-length side-chains (xylose,

galactose) was bound at late, trans cisternae. 5) Antibodies against long XG side-chains, containing xylose, galactose and fucose (M1 and M39) were also found in late cisternae. Of those polysaccharide epitopes which had been previously imaged within the Golgi, cisternal localization results matched earlier findings (41–44).

Following-on from the TEM imaging, the FFE fractions (from R1) were analyzed for the same classes of polysaccharide, using carbohydrate antibody arrays immobilized on nitrocellulose membranes, which has been successfully applied to endomembrane enrichments (45) and post-Golgi compartments (46). Here we were able to probe an expanded number of polysaccharide epitopes compared to TEM due to the high-throughput nature of the array assays. Where possible, antibodies were chosen against epitopes with a known, or likely, sub-Golgi distribution either from previous publications or from Figure 3a. Polysaccharide epitopes were placed into four groups (details in Table S2) with correspondence to the TEM probes. It is notable that the rhamnogalacturonan (RG) class was not covered in the TEM analysis but localized to cis/medial cisternae, as described previously (47), and so was grouped with homogalacturonan and XGs with shorter side-chains. As shown in Figure 3c, the combined FFE profiles from the carbohydrate analysis show distinct distributions for the four epitope groups, with each peaking in the following anode-to-cathode order: complex and medium-branched XG (late), HG, XG with shorter branching and RG (medial), mannans (early) and extensins (very early). Hence, the appearance of the polysaccharide epitopes along the FEE profile has a distinct cisternal bias in the order of trans to medial to cis-Golgi, i.e. going from glycans with more complex or longer to less complex or shorter branching , as the fraction number increases toward the cathode.

Cisternal protein distribution

Next, the overall protein FFE profiles were examined for any evidence of ordering to proteins along the electrophoretic gradient, which might also correspond to different Golgi cisternae. An initial, approximate gauge was obtained by examining the distribution of *N*-glycosylation enzymes where ER or cisternal localization, and hence secretion pathway order, had been established previously (16,48). FUT13, the trans-

Golgi *N*-glycosylation marker, was not present in all replicates, so two alternative biosynthesis enzymes of known trans-Golgi location (49,50) were included. As illustrated in figure 3d, the peak protein abundance was again observed to approximate the late;early;ER sequential order, i.e. with proteins from the medial and trans cisternae more abundant in earlier fractions (closer to the anode). The COPII-associated proteins p24δ2 and p24δ5 were also included for comparison. As anticipated, these profiles were similar to ER and cis-Golgi proteins.

A second, more in-depth protein analysis was conducted using targeted proteomics for proteins previously localized at sub-Golgi resolution (Figure 3e). The notion here was that a higher-sensitivity, but lower-bandwidth technique could be used to validate and complement the high-throughput shotgun proteomics mass spectrometry technique we were using in the main (51,52). The proteins of known localization that were used as sub-Golgi markers for targeted proteomics are listed in Table S2, and include N-glycosylation markers from Figure 3d. Profiles obtained using targeted (Figure 3e) and shotgun proteomics (Figure 3d) were comparable and, again, a cis-medial-trans Golgi trend towards the anode was evident. Together with carbohydrate data, this analysis further corroborated that FFE can separate Golgi cisternae, with earlier cisternae migrating further towards the cathode during separation.

Sub-cluster discrimination

After establishing the general, peak cisternal ordering along the FFE gradient we returned to analysis of the minor sub-Golgi FFE clusters. Following-on from the initial hierarchical clustering of protein abundance profiles, we next generated a more robust set of clusters using a bootstrapping approach, as detailed in the Methods and illustrated in Figure 4a, which randomly omitted 20% of the proteins during repeat hierarchical clustering to generate consensus groups and a measure of uncertainty. This more general, consensus clustering generated clusters numbered 1-8 (Figure 4b). Consistent with the observation that more anodic clusters contained later Golgi proteins, proteins previously localized to the late Golgi (FUT12, XYLT, FUT1, QUA2 – see Figure 3d, e, Table S2) were found in clusters 1 and 2, and proteins previously localized to the early Golgi (GMII, MNS2 – see Figure 3d, e, Table S2) were found in cluster 3 and 4.

Given this, together with the general cisternal separation, we tentatively assigned clusters as follows 1: trans Golgi, 2: medial-Golgi and TGN, 3: cis Golgi, 4:cis-Golgi and ER, 5 & 6:ER, 7 & 8 plasma membrane.

To visualise these clusters on a two-dimensional map, and thus to better illustrate group relationships, PCA was performed on the merged FFE protein profiles (R3-5) using robust clusters 1 – 8 as labels (Figure 3c). Here, Golgi clusters 1, 2 and 3 (trans, medial and cis) formed a somewhat continuous grouping, whilst Golgi cluster 4 was peripheral to the ER group. Cluster 1, the largest Golgi cluster, appeared to be more diffuse at its outer edge but this peripheral group did not obviously correspond to any sub-cluster, so cluster 1 was not further divided. It is notable that two medial-localized *N*-glycosylation enzymes, XYLT and FUT12, were consistently identified in the peripheral region of cluster 1. The dispersed, distal end of cluster 1 might correspond to Golgi residents in a specific trafficking pathway. Although cluster 4 was proximal to the ER cluster, the earlier LOPIT analysis had confirmed that clusters 4 members were resident Golgi proteins. This proximity to the ER suggests a similarity in compartment surface charge, hinting that cluster 4 may be either an intermediate compartment or a Golgi sub-compartment that accepts ER vesicles.

To generate final proteome lists the robust clusters 1-8 were used as labels for training data in a multi-dimensional SVM based classification. This was used to further classify data from R3 – R5, this time considering proteins only detected in single replicates. These additional proteins clustered consistently and so were incorporated into an expanded training set, which was then used in a second round of SVM, to classify merged data from all replicates R1-R5 (See Figure 3d for 2D PCA projection). In the end this yielded compartment proteomes of the following sizes: ER; 181, cis-Golgi; 41, medial Golgi; 56, trans-Golgi; 84 proteins (Table S3).

Golgi cisternae were not expected to differ sufficiently in density to be separable on a density gradient and LOPIT proteome maps were therefore not expected to reflect clustering observed in FFE data. Nevertheless, for comparison, sub-proteomes were plotted onto LOPIT data (Figure 4f). This revealed separate partitioning from the ER and, unexpectedly, some partial separation of Golgi cisternae proteins. The proposed

proteomes largely separated along an ER-cis-medial-trans axis, indicating that classifications from electrophoretic separations were correct.

Validating Golgi cisternae separation

Super-resolution imaging of protein distributions

Next, we validated our observations by testing whether members of the sub-Golgi proteomes showed their proposed in-vivo localizations. Using Structured Illumination Microscopy (53) of transiently-transformed tobacco leaves, we resolved RFP- and GFP-tagged protein pairs for cis/cis, medial/medial, trans/trans, cis/medial, medial/trans, and cis/trans locations (Figure 5a). Proteins were selected based on their functional association with cisternae or relevance to products localized in Figure 4. A visual overview of protein localization is provided in Figure 5a by showing protein localization in individual Golgi stacks. We sampled a large number of Golgi stacks from multiple images (Dataset S1) to generate a statistically robust analysis of protein-pair localization.

To give a measure of the overlap between the locations of the fluorescent proteins, we used a method based upon the distance transform (see Methods) to quantify how coincident the red and green signal intensities were in the Golgi image regions. From the values of the distance transform we devised a simple log-ratio based score to indicate whether the overall distribution of values for the two channels were generally overlapping (positive), partly overlapped (near zero) or separated (negative): examples of this are illustrated in Figure 5b. Results showed that values became more negative (more separated) when combinations were predicted to be more physically distant within the Golgi stack (Figure 5c). Results therefore confirm cis-/medial/trans-Golgi separation using FFE and subsequent compilation of relevant sub-proteomes.

Distribution of protein function across Golgi cisternae

The sub-Golgi proteomes were examined for evidence of functional differences associated with cisternae and were contrasted with the ER and PM. Proteins were grouped by sub-family where possible, given that functional categories such as

‘hemicellullose biosynthesis’, for example, were too broad for the high spatial resolution of Golgi biosynthetic processes. As summarized in Figure 5d for selected groups (see full descriptions in Table S3), specific functions were clearly associated with cisternal sub-proteomes. There was little overlap of typical ER functions (or KDEL motif proteins) with the cis-Golgi and virtually no overlap of typical Golgi functions with the plasma membrane. Prolyl-4-hydroxylases were clearly cis-Golgi associated, as anticipated (54). The GT47 family was enriched in the trans-Golgi, as were glucuronic acid and xylose epimerases. A distinct cis/medial trend was observed in the GAUT and O-fucosyltransferase families.

Bioinformatics analysis of Golgi and sub-Golgi trends

Paralogue TM region sequences

Having established proteomes for the sub-Golgi we sought to identify features common to these sub-compartments that might determine localization. We investigated proteins in our dataset that are close paralogues (i.e. with highly similar amino acid sequences), but which have different cisternal localizations. We observed that the transmembrane (TM) and near-TM regions of the paralogue sequences seemed somewhat variant (Figure 6). Although protein transmembrane regions, because they form simple spanning helices, would be expected to vary somewhat during evolution (not withstanding restraints on hydrophobicity) they are potentially ideal sites for specifying localization given that they can vary without affecting globular domains and are able to respond to a lipid membrane environment. Hence, we investigated the amino acid composition of the TM regions in detail to discern any compartment-specific patterns. As highlighted in Figure 6, an initial casual check on the sequences showed that the paralogues from later cisternae generally had more phenylalanine residues on the exoplasmic/lumenal side of the TM/span and more serines on the exoplasmic side after the TM span.

Compartmental TM region logo plots

To give a more general picture of TM region composition in the cisternae, and because differently localized paralogues are rare, we looked at the overall sequence properties of each localized sub-proteome group. Datasets for single TM-span proteins were

augmented using a similar approach to (11), with only very close homologues selected and TM span edges determined from multiple-alignments using a consistent, hydrophobicity-based informatics procedure (see Methods). We did this for all our localized single-span TM proteins using logo plots for visualization (Figure 7), and where we aligned different sequences according to the cytoplasmic edge or exoplasmic edge of their estimated TM span. This revealed several features that appeared to correlate with progression though either the Golgi stack, or through the entire secretory pathway from ER to PM. From ER to PM there was an increased frequency of Arg/Lys at the cytoplasmic TM boundary. Also, the peak Arg/Lys abundance appears to be broader in the early Golgi compared to the ER. In the Golgi there was increased Ser occurrence at the exoplasmic boundary, although a much weaker, diffuse Ser signal was present in TGN and PM proteins. From the cis- to trans-Golgi, Phe distribution became progressively more biased towards the exoplasmic half of the TM span and Val to the cytoplasmic half, but little change was seen in other hydrophobic residues within Golgi groups. In the PM, Ala, Val, Gly and Ile were predominant in the exoplasmic TM half with Phe and Leu in the cytoplasmic half. Phe frequency was proportionally much lower in the TGN and PM TM-span compared to the ER and Golgi. Hence, overall differences in amino acid distribution likely reflect organelle-specific changes in overall membrane composition and relative differences between the inner and outer membrane leaflets (55).

Intra-protein sequence patterns

Although logo plots of aligned sequences provide a good illustration of amino acid composition, they only present an average picture and are agnostic as to residue correlations within individual sequences. Hence, we additionally analyzed single-span Arabidopsis transmembrane proteins, at the TM spans and ± 15 flanking residues, to look generally for patterns across the (sub-)compartments that were hinted at when inspecting the logo plots together with example sequences.

Firstly, we looked at trends which we would predict from the logo plots by investigating Arg, Lys and Ser residues at TM edges (Figure 8ai). Consistent with the logo plots these showed some abundance differences for Arg/Lys at the cytoplasmic boundary

and Ser at the exoplasmic boundary. However, overall these trends were not especially discriminating for individual cisternae.

Next, we looked at Phe and Ser residues in more detail, given our initial observations on paralogues. Specifically, we measured the asymmetry of Phe composition by comparing the cytoplasmic and exoplasmic halves of the TM span sequences (Figure 8aii). Notably, although overall TM Phe abundance was similar across compartments, Phe was more concentrated in the exoplasmic half of the medial and trans-Golgi TM spans , while the PM, and to some degree the ER, showed the opposite tendency. When looking at Ser abundances (Figure 8aiii) we saw that this increased through the secretory pathway, peaking in the trans-Golgi before dropping a little in the PM. However, a more striking observation was discovered when looking at the presence of three or more adjacent serines (i.e. “SSS” in the sequence) on the exoplasmic side of the TM span; these only seemed to occur in the late Golgi to PM and peaked in the trans-Golgi.

An overview of these results is presented in Figure 8aiv and expressed as a proportion of each sub-proteome, to illustrate the ubiquity of the trends. Overall, although each feature many not be present in all proteins of a given compartment, there is very clearly a fingerprint of characteristics for each. These measures are similar for the cis-Golgi and ER, the TM Phe asymmetry and exoplasmic Ser distinguish later cisternae and cytosolic edge Arg/Lys (i.e. positively charges) are characteristic of trans-Golgi and PM. These features can potentially account for much of the residue intra-Golgi TM protein distribution. However, physical properties like hydrophobicity, exoplasmic and cytoplasmic pI, as we examine next, may also contribute.

TM span properties

When analysing the derived, physical TM span properties it was pertinent to investigate span length, as this is one of very few characteristics associated with increasing membrane thickness in later cisternae (12), although the span-length variety in plant Golgi proteins (16) implies the existence of other factors beyond those specific to protein families (13,14).

As shown in Figure 8b, the span length distributions for the cis-Golgi are similar to those of the ER, and then from the medial Golgi onwards the length tends to increase, on average, through the secretory pathway to the PM. The cytoplasmic pI distributions show analogous trends, albeit with the pI diminishing from the medial Golgi to PM. On the other side of the TM span the exoplasmic pI is somewhat different between the cis-Golgi (lower) and later cisternae (higher), and both are distinct from the ER and PM. The per-residue hydrophobicity (relative to the TM edge) generally reflected the observed trends in TM span length. However, the most notable hydrophobicity differences occurred in the 10- to 15-residue segment flanking the exoplasmic TM boundary (Figure 8c). This increased in the Golgi, from cis- to trans- but was appreciably lower in the TGN and PM. This was accompanied by a decrease in mean exoplasmic residue charge in the late Golgi, which also contrasted with the TGN and PM.

DISCUSSION

This study shows that the secretory pathway can be directionally separated, from the ER to the trans-Golgi. We describe the first proteomic comparison of separated Golgi cisternae and present a series of protein characteristics likely to affect protein location and longevity in different cisternae, along with the most comprehensive Arabidopsis Golgi resident proteome, to date. Separation results were validated by comparing protein and glycan localization in-vivo and post-electrophoresis. Partial separation of cisternae by density gradient centrifugation provided additional independent validation of the cisternal proteomes.

The medial and trans-Golgi are proposed to be the principal sites of polysaccharide synthesis (41) and glycan complexity and length of side-chains is known to increase from cis- to trans-Golgi (3). Our results agree overall but showed considerable levels of polysaccharide synthesis in the early Golgi (Figures 3b & 3c). Consistently fewer gold particles were detected in early compared to late Golgi compartments (Figure 3a), suggesting that polysaccharides are less readily detectable in the early Golgi using immunogold-TEM. The signal from antibodies in the ‘very early’ group was found to persist through Golgi-containing fractions (Figure 3c), even though the LM1 signal was

restricted to the cis-Golgi in TEM images (Figure 3b). Some ‘very early’ antibodies may exhibit some cross reactivity with arabinogalactan side-chains (56), which may be present in the later Golgi. The overall increase in glycosyl-transferase (GT) proteins in the trans-Golgi (Figure 3e) indicates that diversity of glycosylation reactions is greatest in the trans-Golgi.

Functional analysis of cisternal proteomes supported the canonical view that molecular complexity of modified cargo increases through the Golgi and showed that our sub-Golgi categorization accurately reflects biological function. The GAUT family members which have been biochemically characterized are known to synthesize polysaccharide backbones (57), whilst members of the GT47 family and core-2/I-branching beta-1,6-GlcNAc transferase family transfer sugars to peripheral glycan branches (58–62). As shown in Figure 5, the latter two families were found mainly in the trans-Golgi and GAUTs in the cis/medial Golgi. Several of the cis Golgi-localized (Figure 5) P4H enzymes catalyze the first step in *O*-linked glycosylation, and shuttle between the ER and cis-Golgi (54,63), whilst the medial RRA3 (M4 in Figure 5a) catalyzes the subsequent arabinosylation of hydroxyproline (64). Some SAM-dependent methyltransferases have been associated with methyl esterification of substrates synthesized by GAUT1 and GAUT7 (65,66). Consistently, these proteins localized subsequent to GAUT1 and GAUT7 (Figure 5), as did their reaction products (Figure 3). MUR3, a GT47 family member, was located in the trans-Golgi (Figure 5), along with its product, galactosylated xyloglucan (Group 4, Figure 3). Functional insight imparted by our results is demonstrated by analysis of DUF707 proteins, which are suggested to be a GT family (23) but are otherwise unstudied. Of the 11 Arabidopsis family members, we identified 9 in the medial/trans-Golgi cluster in our LOPIT data. In electrophoretic data 3 were identified, all exclusive to the trans-Golgi. Given their trans-Golgi association, and the family size, it seems likely that DUF707s make an important contribution to the diversity of terminal substitutions on glycan chains, possibly relating to the cell wall.

Non-fucosylated xyloglucan epitopes were not observed in the very latest Golgi cisternae (Figure 3a, b, c) but have been recorded in post-Golgi compartments and the

cell wall (46). This suggests that their absence from the very late Golgi was not a consequence of further substitution preventing antibody binding. Possibly, epitopes not being further substituted pass through the very latest cisternae quickly, so are present at low concentrations, though it cannot be ruled out that select cargo may somehow bypass terminal cisternae.

A unique advantage of this study is that hundreds of cargo and resident proteins were tracked simultaneously through the secretory pathway. Profiles of these protein groups indicated distinct trafficking mechanisms; the flat profiles of cargo proteins (Figure S2) were compatible with a uniform, non-selective mechanism of trafficking cargo from the cis- to trans-Golgi, such as cisternal maturation (67). Golgi residents accumulated above cargo abundance levels, which is most straightforwardly explained by recycling of resident proteins directionally opposite to the cargo flow, although anterograde trafficking mechanisms cannot be ruled out. Observations are therefore consistent with the current consensus model of combined cisternal maturation and retrograde vesicular trafficking (67–69). Interestingly, TGN proteins were somewhat more associated with medial than trans-Golgi cisternae (Figure 2d). This could be a consequence of medial Golgi receiving retrograde trafficked material in COPIb vesicles, as recently discussed in (70).

The gradient of increasing electronegativity that appears to exist across the Golgi stack cannot be explained by bulk changes in cytosolic pI of proteins (Figure 8b), so must be attributed to lipid content. Phosphatidyl serine (PS) is an endomembrane-associated monoacidic phospholipid whose concentration at the cytoplasmic leaflet is higher in the Golgi than the ER (71,72) due to the action of flippases (73). Our data indicate that cytoplasmic-leaflet PS concentration increases from cis- to trans-Golgi. In this case, an extremely anodic migration of PM could have been expected, owing to accumulation of cytoplasmic leaflet phosphatidylinositol-4-phosohate (PI(4)P) (72,74). The observed extremely cathodic migration (Figure S2b) was therefore likely due to binding proteins, counterions or most vesicles being in a exoplasmic-face out orientation..

Phe asymmetry in the TM-span, exoplasmic Ser concentration, multiple consecutive Ser, exoplasmic pI and exoplasmic hydrophobicity were convincingly associated with

the later Golgi (Figures.6, 7 & 8). The changes in Phe asymmetry at the TGN and PM (Figure 7) suggests this is an important identifier of Golgi residents. Phe stabilizes membrane proteins by inserting into the bilayer adjacent to ionic lipid:protein interactions (75). Less asymmetric proteins could be progressively excluded if this feature confers stability in the late Golgi luminal environment. The luminal pH of plant secretory compartments decreases from the ER to the TGN, and thereafter increases (76). Total Golgi measurements in earlier studies suggest this feature is not unique to plants (11,77). Exoplasmic Ser could further increase stability in tightly appressed trans-Golgi cisternae by facilitating hydrogen bonding and compact folding through its action as a flexible linker between the TM helix and catalytic domains (11). The increase in Arg/Lys at the cytoplasmic TM boundary from the ER to PM (Figure. 7) may increase protein stability as outer-leaflet concentrations of negatively charged lipids increase throughout the entire secretory pathway. The observation that differences in these sequence features can be detected between differentially localized proteins of very high overall sequence similarity (Figure 6) lends weight to these features being important determining factors in sub-Golgi localization. Recently, Glu at the exoplasmic TM boundary was found to confer cis/medial Golgi localization of GnTI (70). Exoplasmic anchoring of medial protein TM span sequences reveals a prominent Glu at this position in our data, suggesting that multiple Medial Golgi protein are localized in this way. A single, cisternally-specific amino acid at this location was not evident in cis or trans proteomes (Figure 7). If early to late cisternal localization is conferred by a gradient of preference for Ser and Phe, a specific central Golgi signal may add a further level of distinction. Alternatively, this may identify a specific retrieval pathway for medial proteins (70).

At the TGN, most resident proteins must be retained and recycled, and relevant proteins selected for onwards trafficking (78). The drop in exoplasmic pI and hydrophobicity at the TGN (Figure 8c), and loss of exoplasmic Ser (Figure 8a) indicates a sudden change in luminal environment, which could exclude Golgi residents from most TGN regions. Lipid zonation occurs within the TGN (79,80); the decrease in TM Phe bias in TGN proteins indicates that TM span composition may exclude Golgi residents from certain TGN zones. Residue composition appears to play an important role in distinguishing

PM proteins (Figure 7), for example illustrated by the lack of Leu and prominence of Ile toward the exoplasmic TM edge, which is not observed in other membranes. Also, the strong, regular spacing of Gly residues toward the exterior of the TM span may indicate the presence of dimerization sites in these PM proteins (81).

In summary, we have shown the electrophoretic separation of Golgi cisternae is possible and provides a means to determine the order of proteins, and hence functions, within the secretory pathway, and to discriminate residents from cargo. Through this separation we have also uncovered a continuum of differences in transmembrane amino acid sequences across the different Golgi cisternae. Our results provide a framework upon which the precise mechanisms of cisternal localization and longevity can be investigated and will contribute to an understanding how the complex equilibrium of the Golgi is maintained.

METHODS

Preparation of intact-membrane material

Arabidopsis cell suspension culture line (cv. L. erecta) was maintained, homogenized and enriched for endomembranes in a similar manner to (27). For membrane separations, 60 – 80 g fresh weight (FFE separations) or 40 g fresh weight (LOPIT) of 7 day-old cells were protoplasted according to (82) and gently homogenized using 6 strokes of a glass-teflon homogenizer in a 10 mm Na_2HPO_4, 3 mm EDTA, 2 mm dithiothreitol, protease inhibitor tablets (Roche), and 1% dextran 200000 [w/v] buffer (1:2 w/v ratio of fresh cell weight:buffer). The ensuing homogenate was clarified at 3000 *g* for 15 min, then collected on a cushion of 1.4 M sucrose at 100,000 *g* for 1.5 h. The cushion was overlayed with homogenization buffer containing 1.0 M and 0.2 M sucrose and endomembranes were collected at the 1.0/0.2 M interface after centrifugation for 100,000 g for 1.5 h. Each biological replicate (FFE and LOPIT experiemnts) represented a separate preparation of homogenized cell-suspension culture, collected in different weeks, from different inoculations.

Free-Flow Electrophoresis

The electrophoresis was performed using continuous zone electrophoresis-FFE (ZE-FFE) using an FFE System (BD Diagnostics) in the same manner as (27), on five separate biological replicates of endomembrane-enriched samples from Arabidopsis cell-suspension cultures (as above). Separation was by the tangential action of laminar flow and voltage; using 700 V, which resulted in a current of 105–115 mA. The media injection speed was 200 mL/h, and samples at 1500 µL/h. Fractions were collected and assessed for total protein content according to absorbance at 280 nm. Fractions corresponding to the main endomembrane separation zone (See Figure 1) were analyzed using shotgun proteomics (all replicates) and further validated using targeted proteomics (replicate 4) and glycan epitope analysis (replicate 1) where material was available.

Mass spectrometry analysis of replicates 1 and 2

Proteins were reduced, alkylated and digested with trypsin (1:10 w/w) overnight in 50% acentonitrile and 10 mM Tris-HCL, pH 7.5. Peptides were injected onto a Pepmap100 µ-guard column on a Famos Autosampler (both Dionex-LC Packings, Sunnyvale, CA) and washed for 10 min with Buffer A (2% acetonitrile, 0.1% formic acid) flowing at 15 µL/min. Peptides were eluted onto an Acclaim Pepmap100 C18 column (75 µm × 150 mm, 300 nL/min flow rate; Dionex-LC Packings) and into the TripleTOF 5600 via a gradient of 5% buffer B (98% acetonitrile, 0.1% formic acid) increasing B to 35% B over 60 min. B was increased to 90% over 3 min and held for 15 min followed by a ramp back down to 5% B over 3 min where it was held for 15 min to re-equilibrate the column. Peptides were introduced to the mass spectrometer using a Nanospray III source (AB SCIEX) with a nanotip emitter (New Objective, Woburn, MA) in positive-ion mode (2400 V). Data were acquired with Analyst TF 1.5.1 operating in information dependent acquisition (IDA) mode. After a 250 ms scan, the 20 most intense ions (charge states 2–5) within 400–1600 m/z mass range above a threshold of 150 counts were selected for MS/MS analysis. MS/MS spectra were collected using time of flight (TOF) resolution mode: high resolution with the quadrupole set to UNIT resolution and rolling collision energy to optimize fragmentation. MS/MS spectra were scanned from 100–1600 m/z

and were collected for 50 ms. Selected precursor ions were excluded for 16 s following MS/MS acquisition.

Mass spectrometry analysis of replicates 3-5

Proteins were digested as above and resulting peptides were injected on to a Q-Exactive+ (Thermo Fisher Scientific) using a nanoACQUITY UltraPerformance LC system (Waters), incorporating a C_{18} reverse phase column (Waters; 100 µm × 100 mm, 1.7 µm particle, BEH130C18, column temperature 40 °C). Peptides were analysed over a 150- min gradient using Buffer A, 5% Buffer B. Buffer B was increased from 2 to 10% over 2 min, to 40% over 110 min, then to 85% over 1 min, maintained at 85% for 10 min and equilibrated for 14 min with 2% buffer B. Peptides were eluted at a flow rate of 300 nl/min. An MS survey scan was obtained for the m/z range 300–1600. MS/MS spectra were acquired using a top 15 method, where the top 15 ions in the MS spectra were subjected to high- energy collisional dissociation. An isolation mass window of 2.0 m/z was used for the precursor ion selection, and normalized collision energy of 27% was used for fragmentation. A duration of 5 sec was used for the dynamic exclusion. An automatic gain control target of 1,000,000 for MS and 50,000 for MS/MS was used, while maximum IT for MS was 30 ms and MS/MS was 50 ms. The system employed a resolution of 70,000 for MS and 17,500 for MS/MS.

Label-free protein quantitation using the Normalized spectral index (SIn)

Identification annotations were extracted from mzIdentML files. Spectra were clustered using the spectra-clusetr-cli version 1.0.3 (83) a precursor tolerance of 2 m/z and a fragment tolerance of 0.1 m/z. All other settings were left at their defaults. The accuracy of label-free quantitation was improved using the id_transferer_cli tool to transfer identifications to unidentified spectra if these were part of a cluster with >= 5 identified spectra and at least 70% of these spectra identified the same peptide. This approach is comparable to a feature mapping based on precursor m/z and retention time but does not required complex retention time alignment to be performed between the different samples. Proteins were inferred from TAIR10 (84) and the smallest number of proteins required to explain all observed peptides were retained. Peptides that could be

assigned to more than one unambiguously identified protein / protein group were not taken into consideration for label-free quantitation.

Merged FFE profile generation

Fraction-separated spectral count data from different FFE replicates were merged into a single set of pseudo-fraction abundances prior to hierarchical clustering. Merging was achieved by progressive, pairwise aggregation of FFE profiles, using a scheme (see below) that aligns fraction data with the objective of maximizing the correlation between protein abundances in equivalent fractions. Alignment involved an exhaustive search of relative end offsets (and thus linear scaling) to pair-up overlapping/partially overlapping fractions from different experimental replicates. The open-source computer code that performed this operation is available at github.com/tjs23/ms_fraction_merge/.

Fraction align-and-merge procedure:

Data for each replicate, in terms of spectral counts for each protein in each fraction were loaded from CSV files and the later fractions, where total protein count was negligible, were discarded in each case (fractions 43, 50, 47, 37 and 44 respectively for the replicates in this study). Missing abundance values from fractions not harvested after FFE were imputed by performing a linear interpolation of values from the closest fractions either side that were recorded. Each protein's abundance profile in each replicate was then normalized by subtracting its minimum value over all fractions (i.e. base-line correction for those proteins which don’t have a zero-valued fraction) and dividing by the summation of counts; each protein had a fractional abundance profile that summed to 1.0. Each replicate fraction, containing proportional protein abundances, was than normalized by dividing by the fraction's median value, thus centering each fraction irrespective of total protein content. Progressing from the most similar pair of replicates (or replicates combined in a previous round) fraction data was combined by an exhaustive search of relative offset of profile start and end, and hence also width scaling, to find the alignment with the best overall correlation in fraction protein abundances. Fraction offsets, which align the starts and ends of the replicate data, were sampled in the range of +/-5 in steps of 0.5 (original) fraction-widths.

Concomitantly this also sampled fraction width scaling, to shrink or expand the fractions' equivalent range in one data set relative to the other, where intermediate scale values are linearly interpolated. For each combination of start and end offset parameters the similarity between two fractions from two different replicate experiments was calculated as the Pearson correlation in protein abundance (considering proteins common to both replicates) multiplied by the relative width of the overlap between fractions. The width-scaled overlap scores were then summed over all fractions to give an overall replicate-replicate similarity score for each particular combination of offsets. This score is maximized if the replicates are aligned to give equal abundances of each protein in equivalent fractions. The combination of alignment parameters that gave the highest score was then used to merge the replicate data. Merging was achieved by averaging the protein abundances in each pair of equivalent fractions over their region of overlap, and generally resulted in merged pseudo-fractions with non-equal widths (i.e. partial overlap). Where merging was done with data that represents previously combined replicates the protein abundances were scaled proportionality according to the number of original replicates in the combined data. After the first pair of replicates was merged, the next most similar replicate was then merged with the result of the previous merge and the whole procedure was repeated until all replicates had been merged with the rest of the data. After the last merge, new pseudo-fractions were generated by imposing 25 equal width bins on the final data, which averages the protein compositions of differently sized regions that result from the successive rounds of merging. The composition in each bin was simply the average of the protein abundances of the overlapping merge regions weighted according to the width of overlap.

LOPIT analysis and clustering

Organelle separation and fraction collection was performed according to (29) with the following modifications: 20 g fresh weight of cells per gradient were protoplasted and homogenized as described above. Iodixanol was adjusted to the required concentrations using the above homogenization buffers without dextran. Membrane were collected on 25% iodixanol cushions, then adjusted to 25% iodixanol and loaded on to a gradient as described by (29). membranes were then fractionated according to

their density by centrifuging at 100,000 *g* for 8 h in a NVTi65 rotor (Beckman Coulter) using slow braking. Fractions (0.5 ml) were harvested top-down using an Auto Densi-flow collection device (Labconco Corporation, Kansas City, MO).

Fractions were pelleted for 50 mins at 100,000 *g* in an SW55Ti rotor, then resuspended in 25 mM $CaCO_3$ and shaken gently for 30 mins at 4 ˚C before re-pelleting. Membrane in fractions 1, 3, 6, 9, 11, 15, 18 and 20 were sonicated for 15 min in 10 s pulses and assayed for protein content. From each fraction, 80 µg of protein precipitated using chloroform:methanol:water (1:4:3), then resolubilized, and reduced in 50 ul 8 M urea/100 mM HEPES (pH 7.8) containing 0.1% SDS and 7 mM DTT for 2 h (room temperature). IAA was added to a final concentration of 15 mM for 2 h (dark, room temperature). Proteins were precipitated in 6 volumes 80% acetone at -20 °C, then pelleted at 16,000 *g* for 10 min at 8 °C and resuspended in 200 ul 100 mM HEPES pH 8.0. Proteins were digested with sequencing grade trypsin (Promega) for 1 h with a 1:40 enzyme:protein ratio, 37 °C. An additional aliquot of trypsin at 1:40 concentration was added and incubated overnight at 37 °C. Trypsin digests were centrifuged for 10 min at 13,000 *g* to remove any insoluble matter, then reduced to dryness by vacuum centrifugation. TMT 10 plex labelling, peptide de-salting and reverse-phase HPLC were conducted according to (29) but using 100 mM HEPES and ACN instead of TEAB and isopropanol during peptide labelling.

Mass Spectrometry, raw data processing and quantification for LOPIT

All mass spectrometry runs were performed on an Orbitrap Fusion™ Lumos™ Tribrid™ instrument coupled to a Dionex Ultimate™ 3000 RSLCnano system (Thermo Fisher Scientific) with parameters from (20)**.** Raw files were processed with Proteome Discoverer v1.4 (Thermo Fisher Scientific) using the Mascot server v2.3.02 (Matrix Science), searched against the Arabidopsis proteome (canonical sequences, downloaded on 04/02/2017). Precursor and fragment mass tolerances were set to 10 ppm and 0.6 Da, respectively. Trypsin was set as the enzyme of choice and a maximum of 2 missed cleavages were allowed. Static modifications were: carbamidomethyl (C), TMT6plex (N-term) and TMT6plex (K). Dynamic modifications were: oxidation (M), TMT6plex(S), TMT6plex(T). False discovery rate (FDR) was assessed using percolator

and only high confidence peptides were retained. Additional data reduction filters were: peptide rank = 1 and ion score > 25. Quantification at the MS3 level was performed within the Proteome Discoverer workflow using the centroid sum method and an integration tolerance of 2 mmu. Isotope impurity correction factors were applied. Each raw peptide-spectrum match (PSM) reporter intensity was then divided by the sum of all intensities for that PSM (sum normalisation). Protein grouping was carried out according to the minimum parsimony principle and the median of all sum-normalised PSM ratios belonging to each protein group was calculated as the protein group quantitation value. Only proteins with a full reporter ion series were retained.

Machine learning and establishment of resident organelle proteomes using LOPIT data

Data analysis, including PCA, was performed using the R (85) Bioconductor (86) packages MSnbase (87) and pRoloc (30) as described in (28). t-SNE analysis were performed in the R programming environment using Rtsne, with the following parameters: theta=0, perplexity=80, max_iter=800. Supervised machine learning using a SVM classifier with a radial basis function kernel was employed in order to predict the localisation of unlabelled proteins. A training set of organelle markers specific to single subcellular compartments (PM, TGN, Golgi, ER, Peroxisome, Chloroplast, Nucleus, Mitochondria, Cytosol, Vacuole) was compiled by selecting proteins whose combined historical data from confocal microscopy and organelle proteomics, housed in (34) showed a clear majority localization to any one compartment (Table S1). Following the SVM protocol in (28), one hundred rounds of fivefold cross-validation was employed (creating five stratified test/train partitions) to estimate algorithmic performance. This protocol features an additional round of cross-validation on each training partition to optimise the free parameters of the SVM, sigma and cost, via a grid search. Based on the best F1 score (the harmonic mean of precision and recall), for each LOPIT dataset the best sigma and cost were 0.01 and 16, respectively. Previously unclassified proteins with an SVM score greater or equal to the upper quartile value for each compartment were assigned as resident to that compartment if consistently classified in at least 2 of the 4 replicate LOPIT experiments. False assignment rates (FAR) were estimated by

calculating conflicting microscopy data housed in (34) in the new resident organelle proteomes. FAR were between 0.1 and 5% for all locations. The resident Golgi proteome was expanded by lowering the upper quartile threshold until the FAR was 2%. This did not result in the assignment of any new proteins beyond the main Golgi cluster in any replicates, so was deemed an appropriate method for expanding the number of Golgi resident proteins. The final organelle resident proteomes are shown in Table S1.

Hierarchical clustering

Differences between protein abundance profiles were measured as Euclidean distances and grouped using Ward's method for hierarchical clustering, considering the similarity of their merged abundance profiles, across (pseudo-) fractions, from all experimental replicates. Initially, Scientific Python's scipi.hierarchy implementation of Ward's method was used to both create dendrograms and to set the order when plotting the abundance profiles as color matrices (Figures 2 & 4). A threshold was defined after inspection of the dendrogram so that proteins could be split into a useful number of groups (i.e. branches). In each case the threshold was set so that the protein clusters were of roughly equal size and there were at least three predominantly Golgi enriched groups (according to proteomes established by LOPIT). The output order of the initial clustering was also used to set the order of rows and columns in the corresponding correlation matrix shown in Figure S2. This matrix contained the Pearson's correlation coefficient between the (merged) profiles for each pair of proteins.

Immunogold Electron Microscopy

Arabidopsis roots were grown on ½ strength MS media containing 1% sucrose under constant light. Three-day- old root tips were high-pressure frozen, freeze-substituted, embedded, sectioned, and immunolabeled according to (88). Samples were cryofixed in B-type sample holders (Ted Pella) using a Leica HPM-100 high pressure freezer with 1-hexadecene (Sigma) as a cryoprotectant. Samples were freeze-substituted for 5 days at -85 °C in a Leica AFS2 in a solution of 0.25% gluteraldehyde, 0.1% uranyl acetate and 8% 2,2-dimethoxypropane (Sigma) in acetone. Samples were then slowly warmed to room temperature over 2 days, infiltrated with LR White resin (London Resin Company) over 5 days, then resin was polymerized for 36 hours at 70°C. ~70 nm sections were cut

with a DiATOME knife on a Leica UCS ultramicrotome, suspended on nickel grids (Gilder) with 0.3% formvar, blocked with 5% bovine serum albumin in TRIS-buffered saline with detergent (TBST: 10 mM TRIS, 250 mM NaCl, 0.1% w/v tween-20, pH 7.4), and thoroughly washed with TBST before antibody application. Primary antibodies were CCRC-M1, CCRC-M39, CCRC-M87, CCRC-M89 (CarboSource Services www.carbosource.net), LM1, LM15, LM19, and LM21 (PlantProbes http://www.plantprobes.net), all used at 1/10 dilution and applied for 1 hour at room temperature, after which grids were thoroughly washed with TBST. Secondary antibodies were 1/100 goat anti-mouse conjugated to 18 nm gold (Jackson ImmunoResearch 115-215- 146) or 1/100 goat anti-rat conjugated to 18 nm gold (AbCam ab105302), applied for 1 hour at room temperature, after which grids were thoroughly washed with TBST, and then water. Samples were post-stained with 1% (w/v) aqueous uranyl acetate for 8 minutes and Reynolds' lead citrate for 4 minutes. Grids were imaged using a Philips CM120 TEM at 80 kV accelerating voltage coupled to a Gatan multiscan 791 CCD camera. The relative positions of gold particles were determined by measuring the thickness of each Golgi stack from cis to trans, then measuring the distance from the cis-most face for each gold particle

Cis to trans polarity of the Golgi stacks was confidently determined by 1) the cis-to-trans decrease in cisternal lumen width 2) the increase in cisternal lumen electron density (89), and 3) the location of a Golgi-associated TGN (where present). Under our fixation, embedding, and post-staining conditions, the electron density increased up to the trans-most cisternae, usually peaking in the penultimate cisterna. To avoid glancing sections through the margins of Golgi stacks, only Golgis with at least three clearly visible cisterna were imaged.

FFE glycan analysis

The distribution of glycans in replicate 1 after electrophoresis was quantified using carbohydrate microarrays. Here polysaccharides were released from 20 µg of protein by digestion with 4 µg of Proteinase K for 4 h at 37 °C, then dilution in array-jet printing buffer (55.2% glycerol, 44% water, 0.8% Triton X-100) in 0.8% Triton X-100. A 2-fold

dilution series of 4 dilutions was loaded, printed on to nitrocellulose arrays and quantified according to (90), using anti-rat (for LM antibodies) or anti-mouse (for CCRC and BS-400-4 antibodies) secondary antibodies conjugated to alkaline phosphatase (Sigma, Poole, UK). Primary antibodies were sources as above, with the addition of BS-400-4 (Australian Biosupplies (Bundoora, VIC, Australia). For each fraction, extracts equivalent to 0.1 µg or 0.05 µg total protein were probed. Average antibody signal intensities from the dilution series of two technical replicates were normalized to the highest sample value per replicate.

Targeted proteomics

Targeted proteomics of specific proteins was performed on replicate sample 4 by selected reaction monitoring (SRM). This was done on an Agilent 6460QQQ Mass Spectrometer according to (91) using a 25-min method with the following gradient: 95% Buffer A (2% acetonitrile, 0.1% formic acid), 5% Buffer B (98% acetonitrile, 0.1% formic acid). Buffer B increased to 40 % over 17 min, then to 80% B in 30 s, where it was held for 1 min, then ramped back down to 5% in 30 s and equilibrated for 6 min prior to the next injection. Data analysis was performed using Skyline (v2.6) (92). Targets were selected according to the following criteria: confident identification by shotgun mass spectrometry, appreciable increase in signal intensity after enrichment of endomembranes from whole-cell homogenates, appreciable increase in signal intensity after focusing acquisition time around the anticipated retention time. Targets were identified using up to 5 transitions per peptide and at least two peptides per protein (Figure S4). Peptide quantification was achieved by summing the integrated peak areas of two validated SRMs. Peptides were averaged for all proteins associated with sub-compartments. Relative protein abundance was expressed as a percentage of the total for all fractions.

Robust clustering of FFE profiles by bootstrapping

Initial hierarchical clustering of FFE profiles defined eight groups and this number of groups was kept for a second, more robust round of clustering that was less sensitive to the inclusion of specific proteins. The secondary clustering also used Ward’s method

(albeit via the sklearn.custering implementation) but was performed on the rows of the correlation matrix, rather than plain abundance profiles. The robustness of the clustering was assessed by bootstrapping, removing 20% of data each time, and randomly (and independently) re-sampling 120 times. The resulting clusters were colored according to the most similar cluster from the initial (dendrogram) clustering; taking the minimum Euclidean distance between mean correlation profiles, so that the correspondence between the initial and secondary clustering was obvious. The bootstrapping results were then used to estimate the variability in the cluster allocation of each protein (Figure S2). The variability in the assignment of each protein to the clusters was simply measured as the fraction of the bootstrap samples that put the protein in its non-modal cluster.

Machine learning and establishment of resident organelle proteomes using FFE data

SVM was carried out as described earlier for LOPIT datasets but using a best sigma and cost of 0.01 and 16, respectively. Results from the bootstrapped clustering described above provided the training input. Non-Golgi endosomal proteins could be confidently excluded from cis-, medial and trans-Golgi proteomes as all endosomal organelles clustered distinctly in LOPIT analyses. Likewise, this approach was used to distinguish ER and cis-Golgi proteins in clusters 4 and 5. SVM training data and final ER and sub-Golgi proteomes are described in Table S3.

Structured Illumination Microscopy of Golgi stacks

Structured illumination microscopy (SIM) was carried out on the following representative of cis (C), medial (M) and trans (T) proteins: AT2G20810.1 (C1), AT5G47780.1 (C2), AT2G43080.1 (C3), AT1G26850.1 (M1), AT3G62720.1 (M2), AT5G18480.1 (M3), AT1G19360.1 (M4), AT1G74380.1 (T1), AT1G08660.1 (T2), AT4G36890.1 (T4), AT2G35100.1 (T3), AT5G11730.1 (T5). AT1G08660.1, AT2G20810.1, AT5G47780.1, AT5G18480.1, AT3G62720.1, AT5G11730.1, AT2G35100.1, AT1G74380.1, AT1G19360.1 and AT4G36890.1 in pDONR227 were a kind gift from Dr. Berit Ebert. Coding sequences for AT2G43080.1 and AT1G26850.1 were purchased from

914 Arabidopsis Biological Resource Cente (abrc.osu.edu), amplified using Gateway
915 additions for C-terminal tagging using the following gene-specific primers:
916 ATGGCTCCTGCCATGAAG (AT2G43080.1 Fwd), GTAGCTTTTTGCCTCATCC
917 (AT2G43080.1 Rev), ATGGCGTTGAAGTCTAGTTCTG (AT2G26850.1 Fwd),
918 GTGAGTCGAGGTGGAGTTGG (AT2G26850.1 Rev) then recombined into pDONR227.
919 All pDONR227 constructs were recombined into pUBC-GFP_Dest and
920 pUBC_RFP_Dest vectors (93). Sub-Golgi locations of P-UBQ10 driven, C-terminally
921 tagged GFP and RFP fusion proteins were assayed by pairwise comparisons using
922 transient expression in *N. bethamiana* according to (93). Localizations were visualized
923 for the following pairs (see Figure 5a): C1:C2, C2:C3, M1:M2, T3:T2, M2:T1, C2:M2,
924 M2:M3, M4:T4, M3:C3, M1:C1, C1:T5, C1:T3. For each protein pair, three images were
925 taken from at least two leaves. From each image, three regions of >20 Golgi stacks was
926 selected, giving 9 regions per protein pair. Super-resolution images were acquired using
927 a Deltavision OMX 3D-SIM System V3 BLAZE from Applied Precision (a GE Healthcare
928 company) equipped with 3 sCMOS cameras, 405, 488, 592.5 nm diode laser
929 illumination, an Olympus Plan Apo N 60x 1.42 NA oil objective, and standard excitation
930 and emission filter sets. Imaging of each channel was done sequentially using three
931 angles and five phase shifts of the illumination pattern as described in (94). Sections
932 were acquired at 0.125 µm z steps. Raw OMX data was reconstructed and channel
933 registered in SoftWoRx software version 6.5.2 (Applied Precision, a GE Healthcare
934 company). Brightness/contrast was adjusted as necessary using FIJI (95).

935 **Quantification of microscopic image overlap**

936 Analysis of all nine image regions per pair gave a statistically robust analysis of
937 red/green channel overlap. Channel signal overlap was quantified by thresholding
938 intensities to generate regions of interest (ROIs), then summing the distance transform
939 values for one channel’s ROIs within the ROI bounds of the other. Voxelwise nearest-
940 neighbour distances were measured for GFP signal relative to RFP signal using a
941 custom script for Fiji (95) and a custom script in Dataset S1. The latter maps signal
942 volumes using Kapur's maximum entropy thresholding method (96) and measures
943 distances using the exact signed 3D Euclidean distance transform with internal

944 distances set to zero for display on the histogram. The distribution of distances was
945 analyzed by using the log ratio of absolute values and comparing the average positive
946 value with the average (absolute) negative value for each protein pair. Accordingly, log
947 ratios larger than zero indicate overlap, values around zero represent partial overlap
948 and values less than zero indicate separation.

949 **Alignments of similar Golgi proteins sequences from different cisternae**

950 Pairwise sequence alignments were carried out between proteins present in cis-, medial
951 and trans-Golgi proteomes using the nwalign Python module (which implements the
952 Needleman-Wunsch algorithm). Comparisons were ranked according to alignment bit-
953 score and the eight most similar pairs of proteins, representing 3 protein families
954 (GAUTs, GlcNAc transferases, SAM-dependent methyl transferases) are shown in
955 Figure 6.

956 **Identification of TM sequences in localized proteins and close homologues**

957 Analysis of single-span TM protein sequences was performed in a similar manner to
958 previous studies (11,23), albeit with refinements. From the Arabidopsis organelle and
959 sib-Golgi proteome lists, single-span transmembrane proteins were identified by their
960 UniProt database (97) TM span annotation, where it exists, and otherwise by a
961 combination of SignalP 4.0 (98) and TMHMM (99), taking predicted single TM spans
962 and excluding those predicted to be signal peptides. Initial TM span edge positions and
963 cytoplasm-exoplasm chain topology were taken from UniProt, and otherwise from
964 prediction by Phobius (100).

965 Arabidopsis protein sequences were augmented with sequence information from
966 close homologues using BLAST+ (101) searches of the UniProt reference proteomes
967 within the Viridiplantae clade. All searches used an E-value cutoff of 10^{-20}. Overlapping
968 homologue families, from different initial queries, that had common members were
969 separated by allocating each homologue to its most similar query. Resulting family
970 groups all had a single, consistent organelle or sub-compartment annotation that was
971 derived from the Arabidopsis query protein.

Families of sequences were multiply aligned using Clustal Omega (102) with default parameters. TM span edge positions were further refined using the multiple alignment of each homologue family. First, the edges of the TM span (initially taken from UniProt annotations or Phobius) were adjusted within a region of ±5 residues by selecting the point in the alignment with the maximum difference in GES scale hydrophobicity (summed over all proteins in the alignment) between the adjacent five residues on the side of the TM span and the adjacent five residues on the opposite side. Next, the edge positions were trimmed or extended according to the average hydrophobicity over the whole alignment. If the mean hydrophobicity of the next residue exceeded 1.0 KCal/mol (glycine or more hydrophobic) the edge was extended. Similarly, if the mean hydrophobicity of an edge residue was below 1.0 KCal/mol the edge was trimmed. Finally, individual protein adjustments were made, extending or trimming positions for each span sequence. Accordingly, individual TM span edges were trimmed if they ended in a gap or a hydrophilic residue (defined here as Arg, Lys, Asp, Glu, Gln, Asn, His or Ser) or extended if the next residue were suitably hydrophobic (Phe, Met, Ile, Leu, Val, Cys, Trp, Ala, Thr or Gly).

Next, families of proteins were multiply aligned again using Clustal Omega (102) and the following additional checks were made for a comparable TM span, comparing each BLAST+ hit to the query: 1. The length of the protein must not differ by more than 200 residues, 2. There must not be more than four gap insertions in the TM span region 3. The separation from the TM span to the N-terminus must not differ by more than 75 residues, 4. There must be a cursory similarity between span sequences (mean, aligned regional BLOSUM62 score >0.8).

Reduction of protein sequence redundancy

Given that families contain different numbers of protein sequences with different degrees of similarity, each protein was weighted according to its dissimilarity to all other sequences in the whole dataset. Dissimilarity weights for each protein (w_p) were

obtained using a BLAST+ search of each sequence (maximum e-value 10^{-20}) against a database of all the protein sequences and were calculated as:

$$w_p = \frac{1}{\sum_{i=1}^{N_p} \frac{s_i}{m_i^p}}$$

Here, s_i is the BLAST+ bit-score of the aligned high-scoring database hit i (from a total of N_p hits) and m_i^p is the maximum possible bit-score value; the bit score if the query were compared with itself over the same alignment region. Accordingly, a dissimilarity weight is 1.0 if the search only finds itself and approximately $1/N$ if it finds N very similar sequences. This protects against large and/or well conserved protein families having an undue influence on the measurement of general TM span properties.

Protein sequence Logo plots

The frequency of residue occurrence in TM-spans and flanking regions of cisternal proteins and their close homologues was visualized using logo plots. Logo plots were generated by specially written Python scripts (available at github.com/tjs23/logo_plot) after randomly sampling 1000 sequences for each dataset, from position-specific residue abundance probabilities calculated from dissimilarity weighted sequences. The use of dissimilarity weights (as defined above) reduced the effect of redundant sequences, i.e. due to different sized homologous protein families. Different proteins within each sub-group were aligned by anchoring their sequences at the cytoplasmic or exoplasmic edge of the TM-span, prior to generation of logo plots (Figure 7).

Accession Numbers and Data Availability

Electrophoresis proteomics data is deposited to the ProteomeXchange Consortium via the PRIDE partner repository (103) with identifier PXD004596. LOPIT proteomics data is deposited to the ProteomeXchange Consortium via the PRIDE partner repository with identifier PXD009978. SRM data is available from PASSEL, part of PeptideAtlas repository (peptideatlas.org/passel/), accession number PASS00908.

Supplemental Data files

Supplemental Figure 1. t-SNE plots of additional whole-cell Arabidopsis LOPIT experiments.

Supplemental Figure 2. Further details on hierarchical clustering of FFE profiles.

Supplemental Figure 3. Clustering and correlation of FEE profiles from individual replicates 3-5.

Supplemental Figure 4. Comparison of Type II TM protein paralogues with different sub-Golgi classification.

Supplemental Table 1. Resident organelle proteomes from LOPIT experiments after SVM-based classification.

Supplemental Table 2. Additional information for monoclonal antibodies, polysaccharide epitopes and protein targets featured in Figure 3.

Supplemental Table 3. Protein lists for sub-Golgi proteomes.

Supplemental Dataset 1. Complete suite of SIM images used in Figure 5.

https://drive.google.com/open?id=12IjhN17UBVSjfM_x7dyL_RZ1afTM-Zl6

AUTHOR CONTRIBUTIONS

H.T.P. conceptualized the project, prepared all samples and performed all experimental procedures, except where otherwise stated below. H.T.P. and T.J.S. performed data analysis and wrote the manuscript with assistance from K.S.L. T.J.S. performed programming and computational analyses. H.M. collected and analyzed electron microscopy data. J.G. performed label-free quantitation of protein abundances. N.L. and R.B. operated the fluorescence imaging microscope. W.G.T.W. and S.V-M. collected carbohydrate epitope array data. C.J.P. and M.S. oversaw collection of mass spectrometry data. K.S.L. and J.L.H supervised the project and provided laboratory resources. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

H.T.P. expresses thanks to Sean Munro, Nadine Muschalik, Stephen Fry, Janice Millar, Henrik Scheller, Yves Verhertbruggen, Michael Joo and Purbasha Sarkar. Electron microscopy was performed at the Biosciences node of the Melbourne Advanced Microscopy Facility. This project was part of DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. H.T.P. was additionally supported by the Marie Curie fellowship PIEF-GA-2011-301401 (FP7-PEOPLE-2011) and Det Frie Forskningsral-Technologie og Produktion fellowship DFF – 1337-00066. H.E.M. was supported by an EMBO Long-Term Fellowship (EMBO ALTF 1246-2013) and an Australian Research Council Discovery Early Career Researcher Award (DE170100054). J.G. is funded by a grant of the Vienna Science and Technology Fund (WWTF) [project LS11-045]. T.J.S. was supported by the Medical Research Council (MRC file reference number MC_U105178783). Generation of the CCRC series of monoclonal antibodies used in this work was supported by a grant from the NSF Plant Genome Program (DBI-0421683). Distribution of JIM and MAC antibodies used in this work was supported in part by NSF grant DBI-0421683 and RCN 009281.

Bibliography

1. Klute MJ, Melançon P, Dacks JB. Evolution and diversity of the Golgi. Cold Spring Harb Perspect Biol. 2011 Aug 1;3(8):a007849.

2. Strasser R. Plant protein glycosylation. Glycobiology. 2016 Sep;26(9):926–939.

3. Van de Meene AML, Doblin MS, Bacic A. The plant secretory pathway seen through the lens of the cell wall. Protoplasma. 2017 Jan;254(1):75–94.

4. Ito Y, Uemura T, Nakano A. Formation and maintenance of the Golgi apparatus in plant cells. Int Rev Cell Mol Biol. 2014;310:221–287.

5. Brandizzi F, Barlowe C. Organization of the ER-Golgi interface for membrane traffic control. Nat Rev Mol Cell Biol. 2013 Jun;14(6):382–392.

6. Gendre D, Jonsson K, Boutté Y, Bhalerao RP. Journey to the cell surface--the central role of the trans-Golgi network in plants. Protoplasma. 2015 Mar;252(2):385–398.

7. Robinson DG, Pimpl P. Clathrin and post-Golgi trafficking: a very complicated issue. Trends Plant Sci. 2014 Mar;19(3):134–139.

8. Xiang L, Etxeberria E, Van den Ende W. Vacuolar protein sorting mechanisms in plants. FEBS J. 2013 Feb;280(4):979–993.

9. Heard W, Sklenář J, Tomé DFA, Robatzek S, Jones AME. Identification of Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection. Mol Cell Proteomics. 2015 Jul;14(7):1796–1813.

10. Hawes C, Kiviniemi P, Kriechbaumer V. The endoplasmic reticulum: a dynamic and well-connected organelle. J Integr Plant Biol. 2015 Jan;57(1):50–62.

11. Sharpe HJ, Stevens TJ, Munro S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell. 2010 Jul 9;142(1):158–169.

12. Banfield DK. Mechanisms of protein retention in the Golgi. Cold Spring Harb Perspect Biol. 2011 Aug 1;3(8):a005264.

13. Woo CH, Gao C, Yu P, Tu L, Meng Z, Banfield DK, et al. Conserved function of the lysine-based KXD/E motif in Golgi retention for endomembrane proteins among different organisms. Mol Biol Cell. 2015 Nov 15;26(23):4280–4293.

14. Gao C, Cai Y, Wang Y, Kang B-H, Aniento F, Robinson DG, et al. Retention mechanisms for ER and Golgi membrane proteins. Trends Plant Sci. 2014 Aug;19(8):508–515.

15. Saint-Jore-Dupas C, Nebenführ A, Boulaflous A, Follet-Gueye M-L, Plasson C, Hawes C, et al. Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. Plant Cell. 2006 Nov 30;18(11):3182–3200.

16. Schoberer J, Strasser R. Sub-compartmental organization of Golgi-resident N-glycan processing enzymes in plants. Mol Plant. 2011 Mar;4(2):220–228.

17. Tie HC, Mahajan D, Chen B, Cheng L, VanDongen AMJ, Lu L. A novel imaging method for quantitative Golgi localization reveals differential intra-Golgi trafficking of secretory cargoes. Mol Biol Cell. 2016 Mar 1;27(5):848–861.

18. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlén M, et al. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods. 2013 Apr;10(4):315–323.

19. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, et al. Quantitative proteomics analysis of the secretory pathway. Cell. 2006 Dec 15;127(6):1265–1281.

20. Mulvey CM, Breckels LM, Geladaki A, Britovšek NK, Nightingale DJH, Christoforou A, et al. Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. Nat Protoc. 2017 Jun;12(6):1110–1135.

21. Groen AJ, Sancho-Andrés G, Breckels LM, Gatto L, Aniento F, Lilley KS. Identification of trans-golgi network proteins in Arabidopsis thaliana root tissue. J Proteome Res. 2014 Feb 7;13(2):763–776.

22. Dunkley TPJ, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, et al. Mapping the Arabidopsis organelle proteome. Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518–6523.

23. Nikolovski N, Rubtsov D, Segura MP, Miles GP, Stevens TJ, Dunkley TPJ, et al. Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. Plant Physiol. 2012 Oct;160(2):1037–1051.

24. Islinger M, Eckerskorn C, Völkl A. Free-flow electrophoresis in the proteomic era: a technique in flux. Electrophoresis. 2010 Jun;31(11):1754–1763.

25. Barkla BJ, Vera-Estrella R, Pantoja O. Enhanced separation of membranes during free flow zonal electrophoresis in plants. Anal Chem. 2007 Jul 15;79(14):5181–5187.

26. Morré DJ, Mollenhauer HH, editors. Isolation and Subfractionation. The Golgi Apparatus. New York, NY: Springer New York; 2009. p. 39–61.

27. Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, et al. Isolation and proteomic characterization of the Arabidopsis Golgi defines functional and novel components involved in plant cell wall biosynthesis. Plant Physiol. 2012 May;159(1):12–26.

28. Breckels LM, Mulvey CM, Lilley KS, Gatto L. A Bioconductor workflow for processing and analysing spatial proteomics data [version 1; peer review: 1 approved, 1 approved with reservations]. F1000Res. 2016 Dec 28;5.

29. Christoforou A, Mulvey CM, Breckels LM, Geladaki A, Hurrell T, Hayward PC, et al. A draft map of the mouse pluripotent stem cell spatial proteome. Nat Commun. 2016 Jan 12;7:8992.

30. Gatto L, Breckels LM, Wieczorek S, Burger T, Lilley KS. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. Bioinformatics. 2014 May 1;30(9):1322–1324.

31. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science. 2017 May 26;356(6340).

32. Breckels LM, Gatto L, Christoforou A, Groen AJ, Lilley KS, Trotter MWB. The effect of organelle discovery upon sub-cellular protein localisation. J Proteomics. 2013 Aug 2;88:129–140.

33. Breckels LM, Holden SB, Wojnar D, Mulvey CM, Christoforou A, Groen A, et al. Learning from Heterogeneous Data Sources: An Application in Spatial Proteomics. PLoS Comput Biol. 2016 May 13;12(5):e1004920.

34. Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. Nucleic Acids Res. 2017 Jan 4;45(D1):D1064–D1074.

35. Van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res [Internet]. 2008 Nov;9:2579–2605. Available from: http://www.jmlr.org/papers/v9/vandermaaten08a.html

36. Hooper CM, Stevens TJ, Saukkonen A, Castleden IR, Singh P, Mann GW, et al. Multiple marker abundance profiling: combining selected reaction monitoring and data-dependent acquisition for rapid estimation of organelle abundance in subcellular samples. Plant J. 2017 Dec;92(6):1202–1217.

37. Eubel H, Lee CP, Kuo J, Meyer EH, Taylor NL, Millar AH. Free-flow electrophoresis for purification of plant mitochondria by surface charge. Plant J. 2007 Nov;52(3):583–594.

38. De Michele R, McFarlane HE, Parsons HT, Meents MJ, Lao J, González Fernández-Niño SM, et al. Free-Flow Electrophoresis of Plasma Membrane Vesicles Enriched by Two-Phase Partitioning Enhances the Quality of the

Proteome from Arabidopsis Seedlings. J Proteome Res. 2016 Mar 4;15(3):900–913.

39. Zielinska DF, Gnad F, Schropp K, Wiśniewski JR, Mann M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. Mol Cell. 2012 May 25;46(4):542–548.

40. Yeats TH, Bacic A, Johnson KL. Plant glycosylphosphatidylinositol (GPI) anchored proteins at the plasma membrane-cell wall nexus. J Integr Plant Biol. 2018 Apr 18;60(8):649–669.

41. Driouich A, Follet-Gueye M-L, Bernard S, Kousar S, Chevalier L, Vicré-Gibouin M, et al. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Front Plant Sci. 2012 Apr 30;3:79.

42. Viotti C, Bubeck J, Stierhof Y-D, Krebs M, Langhans M, van den Berg W, et al. Endocytic and secretory traffic in Arabidopsis merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle. Plant Cell. 2010 Apr 30;22(4):1344–1357.

43. Marcus SE, Verhertbruggen Y, Hervé C, Ordaz-Ortiz JJ, Farkas V, Pedersen HL, et al. Pectic homogalacturonan masks abundant sets of xyloglucan epitopes in plant cell walls. BMC Plant Biol. 2008 May 22;8:60.

44. Smallwood M, Beven A, Donovan N, Neill SJ, Peart J, Roberts K, et al. Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex. Plant J. 1994 Feb;5(2):237–246.

45. Okekeogbu IO, Pattathil S, González Fernández-Niño SM, Aryal UK, Penning BW, Lao J, et al. Glycome and proteome components of Golgi membranes are common between two angiosperms with distinct cell wall structures. Plant Cell. 2019 Mar 26;

46. Wilkop T, Pattathil S, Ren G, Davis DJ, Bao W, Duan D, et al. A Hybrid Approach Enabling Large-Scale Glycomic Analysis of Post-Golgi Vesicles Reveals a Transport Route for Polysaccharides. Plant Cell. 2019 Mar;31(3):627–644.

47. Ralet M-C, Tranquet O, Poulain D, Moïse A, Guillon F. Monoclonal antibodies to rhamnogalacturonan I backbone. Planta. 2010 May;231(6):1373–1383.

48. Nilsson T, Au CE, Bergeron JJM. Sorting out glycosylation enzymes in the Golgi apparatus. FEBS Lett. 2009 Dec 3;583(23):3764–3769.

49. Gao P, Xin Z, Zheng Z-L. The OSU1/QUA2/TSD2-encoded putative methyltransferase is a critical modulator of carbon and nitrogen nutrient balance response in Arabidopsis. PLoS ONE. 2008 Jan 2;3(1):e1387.

50. Chevalier L, Bernard S, Ramdani Y, Lamour R, Bardor M, Lerouge P, et al. Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. Plant J. 2010 Dec;64(6):977–989.

51. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell. 2009 Aug 21;138(4):795–806.

52. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat Methods. 2010 Jan;7(1):43–46.

53. Heintzmann R, Huser T. Super-Resolution Structured Illumination Microscopy. Chem Rev. 2017 Dec 13;117(23):13890–13908.

54. Yuasa K, Toyooka K, Fukuda H, Matsuoka K. Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. Plant J. 2005 Jan;41(1):81–94.

55. Xu H, Su W, Cai M, Jiang J, Zeng X, Wang H. The asymmetrical structure of Golgi apparatus membranes revealed by in situ atomic force microscope. PLoS ONE. 2013 Apr 16;8(4):e61596.

56. Pattathil S, Avci U, Baldwin D, Swennes AG, McGill JA, Popper Z, et al. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. Plant Physiol. 2010 Jun;153(2):514–525.

57. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, et al. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc Natl Acad Sci U S A. 2011 Dec 13;108(50):20225–20230.

58. Iwai H, Hokura A, Oishi M, Chida H, Ishii T, Sakai S, et al. The gene responsible for borate cross-linking of pectin Rhamnogalacturonan-II is required for plant reproductive tissue development and fertilization. Proc Natl Acad Sci U S A. 2006 Oct 31;103(44):16592–16597.

59. Jensen JK, Sørensen SO, Harholt J, Geshi N, Sakuragi Y, Møller I, et al. Identification of a xylogalacturonan xylosyltransferase involved in pectin biosynthesis in Arabidopsis. Plant Cell. 2008 May 6;20(5):1289–1302.

60. Zhong R, Peña MJ, Zhou G-K, Nairn CJ, Wood-Jones A, Richardson EA, et al. Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell. 2005 Dec;17(12):3390–3408.

61. Harholt J, Jensen JK, Verhertbruggen Y, Søgaard C, Bernard S, Nafisi M, et al. ARAD proteins associated with pectic Arabinan biosynthesis form complexes when transiently overexpressed in planta. Planta. 2012 Jul;236(1):115–128.

62. Knoch E, Dilokpimol A, Tryfona T, Poulsen CP, Xiong G, Harholt J, et al. A β-glucuronosyltransferase from Arabidopsis thaliana involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. Plant J. 2013 Dec;76(6):1016–1029.

63. Velasquez SM, Ricardi MM, Dorosz JG, Fernandez PV, Nadra AD, Pol-Fachin L, et al. O-glycosylated cell wall proteins are essential in root hair growth. Science. 2011 Jun 17;332(6036):1401–1403.

64. Chen Y, Dong W, Tan L, Held MA, Kieliszewski MJ. Arabinosylation Plays a Crucial Role in Extensin Cross-linking In Vitro. Biochemistry insights. 2015 Sep 20;8(Supple 2):1–13.

65. Miao Y, Li H-Y, Shen J, Wang J, Jiang L. QUASIMODO 3 (QUA3) is a putative homogalacturonan methyltransferase regulating cell wall biosynthesis in Arabidopsis suspension-cultured cells. J Exp Bot. 2011 Oct;62(14):5063–5078.

66. Krupková E, Immerzeel P, Pauly M, Schmülling T. The TUMOROUS SHOOT DEVELOPMENT2 gene of Arabidopsis encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J. 2007 May;50(4):735–750.

67. Luini A. A brief history of the cisternal progression-maturation model. Cell Logist. 2011 Jan;1(1):6–11.

68. Glick BS, Luini A. Models for Golgi traffic: a critical assessment. Cold Spring Harb Perspect Biol. 2011 Nov 1;3(11):a005215.

69. Donohoe BS, Kang B-H, Gerl MJ, Gergely ZR, McMichael CM, Bednarek SY, et al. Cis-Golgi cisternal assembly and biosynthetic activation occur sequentially in plants and algae. Traffic. 2013 May;14(5):551–567.

70. Schoberer J, Liebminger E, Vavra U, Veit C, Grünwald-Gruber C, Altmann F, et al. Golgi localization of GnTI requires a polar amino acid residue within its transmembrane domain. Plant Physiol. 2019 Apr 10;

71. Leventis PA, Grinstein S. The distribution and function of phosphatidylserine in cellular membranes. Annu Rev Biophys. 2010;39:407–427.

72. Simon MLA, Platre MP, Assil S, van Wijk R, Chen WY, Chory J, et al. A multi-colour/multi-affinity marker set to visualize phosphoinositide dynamics in Arabidopsis. Plant J. 2014 Jan;77(2):322–337.

73. Poulsen LR, López-Marqués RL, McDowell SC, Okkeri J, Licht D, Schulz A, et al. The Arabidopsis P4-ATPase ALA3 localizes to the golgi and requires a beta-subunit to function in lipid translocation and secretory vesicle formation. Plant Cell. 2008 Mar 14;20(3):658–676.

74. Simon MLA, Platre MP, Marquès-Bueno MM, Armengot L, Stanislas T, Bayle V, et al. A PtdIns(4)P-driven electrostatic field controls cell membrane identity and signalling in plants. Nature Plants. 2016 Jun 20;2:16089.

75. Bogdanov M, Dowhan W, Vitrac H. Lipids and topological rules governing membrane protein assembly. Biochim Biophys Acta. 2014 Aug;1843(8):1475–1488.

76. Martinière A, Bassil E, Jublanc E, Alcon C, Reguera M, Sentenac H, et al. In vivo intracellular pH measurements in tobacco and Arabidopsis reveal an unexpected pH gradient in the endomembrane system. Plant Cell. 2013 Oct 8;25(10):4028–4043.

77. Quiroga R, Trenchi A, González Montoro A, Valdez Taubas J, Maccioni HJF. Short transmembrane domains with high-volume exoplasmic halves determine retention of Type II membrane proteins in the Golgi complex. J Cell Sci. 2013 Dec 1;126(Pt 23):5344–5349.

78. Guo Y, Sirkis DW, Schekman R. Protein sorting at the trans-Golgi network. Annu Rev Cell Dev Biol. 2014 Aug 18;30:169–206.

79. Surma MA, Klose C, Simons K. Lipid-dependent protein sorting at the trans-Golgi network. Biochim Biophys Acta. 2012 Aug;1821(8):1059–1067.

80. Wattelet-Boyer V, Brocard L, Jonsson K, Esnay N, Joubès J, Domergue F, et al. Enrichment of hydroxylated C24- and C26-acyl-chain sphingolipids mediates

PIN2 apical sorting at trans-Golgi network subdomains. Nat Commun. 2016 Sep 29;7:12788.

81. Teese MG, Langosch D. Role of gxxxg motifs in transmembrane domain interactions. Biochemistry. 2015 Aug 25;54(33):5125–5135.

82. Eubel H, Meyer EH, Taylor NL, Bussell JD, O'Toole N, Heazlewood JL, et al. Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of Arabidopsis cell culture peroxisomes. Plant Physiol. 2008 Dec;148(4):1809–1829.

83. Griss J, Perez-Riverol Y, Lewis S, Tabb DL, Dianes JA, Del-Toro N, et al. Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. Nat Methods. 2016 Aug;13(8):651–656.

84. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015 Aug 4;53(8):474–485.

85. Team R. R: A language and environment for statistical computing. 2013; Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.5851&rep=rep1&type=pdf

86. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004 Sep 15;5(10):R80.

87. Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics. 2012 Jan 15;28(2):288–289.

88. McFarlane HE, Young RE, Wasteneys GO, Samuels AL. Cortical microtubules mark the mucilage secretion domain of the plasma membrane in Arabidopsis seed coat cells. Planta. 2008 May;227(6):1363–1375.

89. Staehelin LA, Giddings TH, Kiss JZ, Sack FD. Macromolecular differentiation of Golgi stacks in root tips of Arabidopsis and Nicotiana seedlings as visualized in high pressure frozen and freeze-substituted samples. Protoplasma. 1990;157(1-3):75–91.

90. Pedersen HL, Fangel JU, McCleary B, Ruzanski C, Rydahl MG, Ralet M-C, et al. Versatile high resolution oligosaccharide microarrays for plant glycobiology and cell wall research. J Biol Chem. 2012 Nov 16;287(47):39429–39438.

91. Batth TS, Singh P, Ramakrishnan VR, Sousa MML, Chan LJG, Tran HM, et al. A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. Metab Eng. 2014 Nov;26:48–56.

92. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010 Apr 1;26(7):966–968.

93. Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K, Blatt MR. A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J. 2010 Oct;64(2):355–365.

94. Gustafsson MGL, Shao L, Carlton PM, Wang CJR, Golubovskaya IN, Cande WZ, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008 Jun;94(12):4957–4970.

95. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jun 28;9(7):676–682.

96. Kapur JN, Sahoo PK, Wong AKC. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing. 1985 Mar;29(3):273–285.

97. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan;43(Database issue):D204–12.

98. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011 Sep 29;8(10):785–786.

99. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567–580.

100. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004 May 14;338(5):1027–1036.

101. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

102. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol. 2014;1079:105–116.

103. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 2016 Jan 4;44(D1):D447–56.

104. Ramsak Ž, Baebler Š, Rotter A, Korbar M, Mozetic I, Usadel B, et al. GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology. Nucleic Acids Res. 2014 Jan;42(Database issue):D1167–75.

FIGURE AND TABLE LEGENDS

Figure 1. Schematic overview of electrophoretic separation profile analysis of endomembrane proteins.

a. Samples from Arabidopsis cell-suspension cultures, enriched in intact endomembranes, were separated by voltage under laminar flow, i.e. using free-flow electrophoresis (FFE). This provided gentle separation of membrane-bound compartments, according to their surface charges, and resulted in 96 separately collected fractions, ordered along the voltage axis.

b. Total protein content of FFE fractions was determined via absorption at 280 nm to identify the range of fractions with major endomembrane protein enrichment. These and adjacent fractions were then taken forward for more detailed analysis. Non-membrane components from the samples peaked in early fractions outside this range.

c. Endomembrane fractions were primarily investigated using shotgun proteomics, to measure the relative amounts of the different proteins contained therein. Here proteins were identified via the mass fingerprints of trypic digest peptides searched against the most recent Arabidopsis proteome using MASCOT software.

d. Average FFE abundance profiles for resident proteins from Golgi, ER and other organelles, using independent sub-cellular localizations derived from LOPIT analysis (see Table S1). Protein abundance values from multiple replicate FFE runs, in the form of MS spectral intensities, were combined (see Methods for the fraction matching procedure), generating 25 merged, consensus endomembrane fractions. Combined data is shown for totals of 200 ER, 204 Golgi and 1290 other organelle proteins.

Figure 2. Primary determination of organelle sub-proteomes.

a. PCA analysis of a single LOPIT experiment: protein abundance profiles from density-based separation are presented by projection onto their two principle, orthogonal axes, representing most inter-protein variance. Each point represents a single protein, which is colored according to its organelle classification. Organelle clusters were distinguished using multi-class SVM on complete abundance profiles and used existing annotations for classification (see Methods).

b. Presentation of the same LOPIT data and classifications shown in **a.**, presented as a two-dimensional t-SNE plot. This visualization attempts to preserve the proximity of

similar profiles, and the separation of distinct profiles, over all data dimensions (whole profiles). This is unlike PCA which shows (dis)similarity along the selected projection axes.

c. Average FFE profiles, across 25 merged fractions from replicates R3-5 are shown for organelle groups classified using LOPIT data. Plotted values represent the mean abundance for each fraction in each organelle class, from per-protein normalized profiles (see Methods). Error bars represent the standard error of the mean. Data is shown separately for the ER and Golgi (upper plot), which peak as a class in central fractions, and the distinct profiles for other organelles/compartments (lower plot). ER and Golgi proteomes have been sub-divided as either those belonging to the initial organelle markers or those newly classified as organelle residents, demonstrating the accuracy with which new residents were assigned.

d. Hierarchical clustering of secretory (ER, Golgi TGN and PM) protein FFE profiles. Merged abundance profiles from proteins identified in high-quality replicates R3-5 were clustered using Ward's method and presented as a dendrogram with the corresponding, underlying abundance profiles shown beneath as a color density plot, together with primary organelle classifications derived from LOPIT. The three major clusters that separated profiles generally into Golgi/TGN, ER and PM were further separated into eight smaller clusters, labelled A-H. Here a threshold was chosen so that each major ER and Golgi cluster contained three minor clusters.

Figure 3. Establishing characteristics of early and late Golgi FFE profiles.

a. Example negative-stain TEM images showing the in-vivo distributions of several glycan epitopes, with varying structural complexity, across the Golgi stack. Glycans were localized using monoclonal antibodies liked to gold particles. All stacks are depicted with cis at the bottom and trans as the top, is indicated.

b. Violin plots showing the overall data from the immuno-gold TEM localization of glycan epitopes, as illustrated in **a**. The relative Golgi stack positions of gold particles represent the fraction of the particle distance to the outer cis face as a proportion of the total cis-trans thickness.

c. FFE abundance profiles for four classes of glycan epitope, with varying structural complexity. Class members and epitope structures are detailed in Table S2. Data is shown for detergent-extracted samples from FFE replicate R1 that were printed onto nitrocellulose microarrays and probed via alkaline phosphatase linked monoclonal antibodies. Error bars show SEM for n=3 antibodies (Group 4), n=9 (Group 3), n=2 (Group2) and n=5 (Group 1).

d. Exemplar FFE protein abundance profiles, as detected by high-throughput shotgun proteomics. Example proteins were selected on the basis of previously established sub-Golgi, ER and transitional ER-Golgi localization relating to well-known biomolecular functions in the secretory pathway.

e. FEE abundance profiles of selected proteins detected via high-sensitivity, targeted proteomics. Proteins (see Table S2) were chosen given an established function and localization specific to Golgi cisternae or the ER. Two independent peptides per protein were measured for n=7 (ER), n=1 (cis-Golgi), n=5 (medial Golgi) and n=3 (trans-Golgi) proteins. Solid lines indicate mean abundance over all proteins in the class and error bars represent SEM.

Figure 4. Classification of sub-Golgi compartments

a. Robust clustering of secretory protein FFE profiles via bootstrapping. Abundance profiles (second from top) were re-clustered using Ward's method 120 times, each time omitting 20% of the proteins. The resulting clusters were assigned to the corresponding initial clusters A-H (see Figure 2) by similarity to the cluster medioids. These clusters are shown as a color map (third panel) where each row corresponds to a different, random subset of proteins, and is presented in the initial hierarchical cluster column order (as used in Figure 2). The robust, consensus clusters (lower panel) were defined as the most common cluster identity for each protein over all the bootstrap trials.

b. FFE profiles for each of the eight consensus groups were separately re-clustered (Ward's method) to clearly visualize profile characteristics of each group. The groups were re-labelled 1-8 to discriminate them from the initial clusters A-H, which have (slightly) different memberships. These were then used for tentative assignment of

particular groups (1-4) to sub-Golgi compartments using trends presented in Figure 3. Abundance profiles are presented as a color density map, as in **a**, but in a new intra-group order.

c. Merged FFE profile data, for proteins present in replicates R3-5, plotted as a two-dimensional PCA projection and labelled according to the bootstrap consensus clusters 1-8, as illustrated in **b**.

d. Merged FFE profile data for all secretory proteins detected in any of the replicates R1-5, presented as a two-dimensional PCA projection. Multi-class SVM was used to classify proteins (on whole FFE profiles, not the 2-D map) into three sub-Golgi groups and an ER group. The group labels used in the classification came from LOPIT to provide distinction between resident ER and Golgi proteins (and to exclude TGN ones), given that profiles overlap, to a degree, in the FFE data, but not in the LOPIT data. The consensus FFE sub-clusters (as in **c**) were then used to classify the three sub-Golgi groups from among the larger Golgi proteome. Consensus sub-clusters and final proteomes are detailed in Table S3.

e. Re-presentation of a section of the LOPIT PCA map shown in Figure 2a, now colored according to ER and sub-Golgi classes presented in **d**.

f. Re-presentation of a section of the 2D t-SNE map shown in Figure 2b, now colored according to ER and sub-Golgi classes presented in **d**.

Figure 5. Validation of sub-Golgi protein localization

a. Example images of structured illumination microscopy (SIM) of validatory protein pairs, representative of cis (C), medial (M) and trans (T) Golgi sub-localizations. Sub-Golgi locations of P_{UBQ10} driven, C-terminally tagged GFP and RFP fusion proteins (93) were assayed to provide pairwise comparisons by using transient expression in *N. bethamiana*. For each protein pair, localization data were collected from nine regions (Dataset S1), incorporating three image stacks from at least two leaves per plant. Localizations were visualized in a single Golgi body from each of the three image stacks. The gene identifiers for the proteins were: AT2G20810.1 (C1), AT5G47780.1

(C2), AT2G43080.1 (C3), AT1G26850.1 (M1), AT3G62720.1 (M2), AT5G18480.1 (M3), AT1G19360.1 (M4), AT1G74380.1 (T1), AT1G08660.1 (T2), AT4G36890.1 (T4), AT2G35100.1 (T3), AT5G11730.1 (T5). Scale bars = 400 nm.

b. Three example histograms showing the distribution of distance transform values for image regions containing multiple Golgi stacks with spatially overlapping (top), partly overlapping (middle) and somewhat separate (bottom) labelled protein pairs, i.e. from red/green fluorescence microscopy illustrated in **a**. Channel signal overlap was quantified by thresholding intensities to generate regions of interest (ROIs), then summing the distance transform values for one channel's ROIs within the ROI bounds of the other. Here, negative values indicate greater separation and positive values indicate overlap.

c. The distribution of red/green channel overlap scores, over multiple image regions (n=9), for the validatory protein pairs shown in **a**, arranged in modal order. Overlap scores were calculated for each image region as the $\log_2$ ratio of mean absolute values either side of zero distance (see blue and orange regions in **b**), with positive values indicating more overlap. Image regions are given in Dataset S1.

d. Occurrence of proteins families and functional annotation in the secretory and sub-Golgi proteomes. Using ER, TGN and PM localizations derived from LOPIT data and sub-Golgi localizations from FFE (see Table S3), proteins were grouped variously according to family, MapMan (104) functional categorization and possession of the K/H/RDEL ER-retrieval motif. Groups with at least 5 members are presented here.

Figure 6: Comparison of Type II TM protein paralogues with different sub-Golgi classification.

Alignments are shown for pairs of similar, homologous proteins from Arabidopsis which have different sub-Golgi localisations. TM-span regions are idicated in bold. The blue Arg/Lys at the cytoplasmic edge highlight the start of the TM span. Phe residues are colouredeither pink or cyan to indicate relative position in the TM span. Within 15 residues of the exoplasmic TM edge Ser residues are coloured yellow and three consecutive Ser are red.

Figure 7. Transmembrane amino acid composition in sub-Golgi and secretory compartments.

Logo plots of single-span TM proteins from secretory and sub-Golgi proteomes indicating the relative abundance of amino acids at and around aligned TM spans.

Data is shown for the Arabidopsis proteins localized by LOPIT and FFE and their very close homologues. Different sequences were aligned at either the cytoplasmic (left column) or exoplasmic/luminal (right column) edge of the hydrophobic TM spans. (See Methods for details of gathering homologues and aligning TM sequences). The different amino acids are color coded according to their physiochemical properties, as indicated in the color key (bottom). Logo plots were generated after randomly sampling 1000 sequences for each dataset, from position-specific residue abundance probabilities calculated from dissimilarity weighted sequences. This was done to reduce the bias caused by the different sizes of protein families, i.e. which are informatically somewhat redundant.

Figure 8. Comparison of protein sequence features in organelle and sub-organelle proteomes.

a. Abundance of sequence features at and around the TM spans of single-span proteins in the secretory and sub-Golgi proteomes. Data is shown for 63 ER, 23 cis-, 37 medial and 54 trans-Golgi proteins, and 108 PM transmembrane proteins. **i.** The relative abundance of lysine or arginine at the cytoplasmic TM edge and serine at the exoplasmic/luminal edge. Values were normalized relative to the maximum observation. **ii.** Overall TM phenylalanine content, as a proportion of TM span length, and the cytoplasmic-exoplasmic asymmetry of TM phenylaniline; asymmetry was calculated as the difference in the abundance between the two halves of each TM span. **iii.** The relative abundance of Serine and presence of three or more consecutive Serines in the 15-residue exoplasmic region immediately flanking the TM spans. Values were normalized relative to the maximum observation. **iv.** An overview of the results presented in **i-iii**, but shown as a proportion of each sub-cellular proteome. Here phenylalanine asymmetry corresponded to positive values presented in **ii** and high

serine content corresponded to a count of at least 5 in the 15 flanking exoplasmic residues. For panels **i-iii** bar heights are mean values and errors represent the standard error in the mean.

b. Distributions TM span properties for different sub-proteome groups. Datasets for localized single-span TM proteins from Arabidopsis were expanded through close homology searches (as used in Figure 7), where sequence contributions were weighted by dissimilarity and TM-span were edges defined, as detailed in the Methods. TM-span length (top), pI of the entire cytoplasmic region (middle) and pI of the entire exoplasmic region (bottom) are shown as violin plots for different secretory and sub-Golgi compartments (defined by LOPIT and FFE respectively).

c. Line plots of per-position TM hydrophobicity (top) and mean residue charge (bottom) for localized Arabidopsis and homologue over TM hydrophobic core and flanking regions (as in Figure 7). TM spans were anchored at their exoplasmic boundary. Plotted values represent the means at each TM aligned position, over different, dissimilarity-weighted proteins. Error bars represent the standard error in the mean.

Figure 1

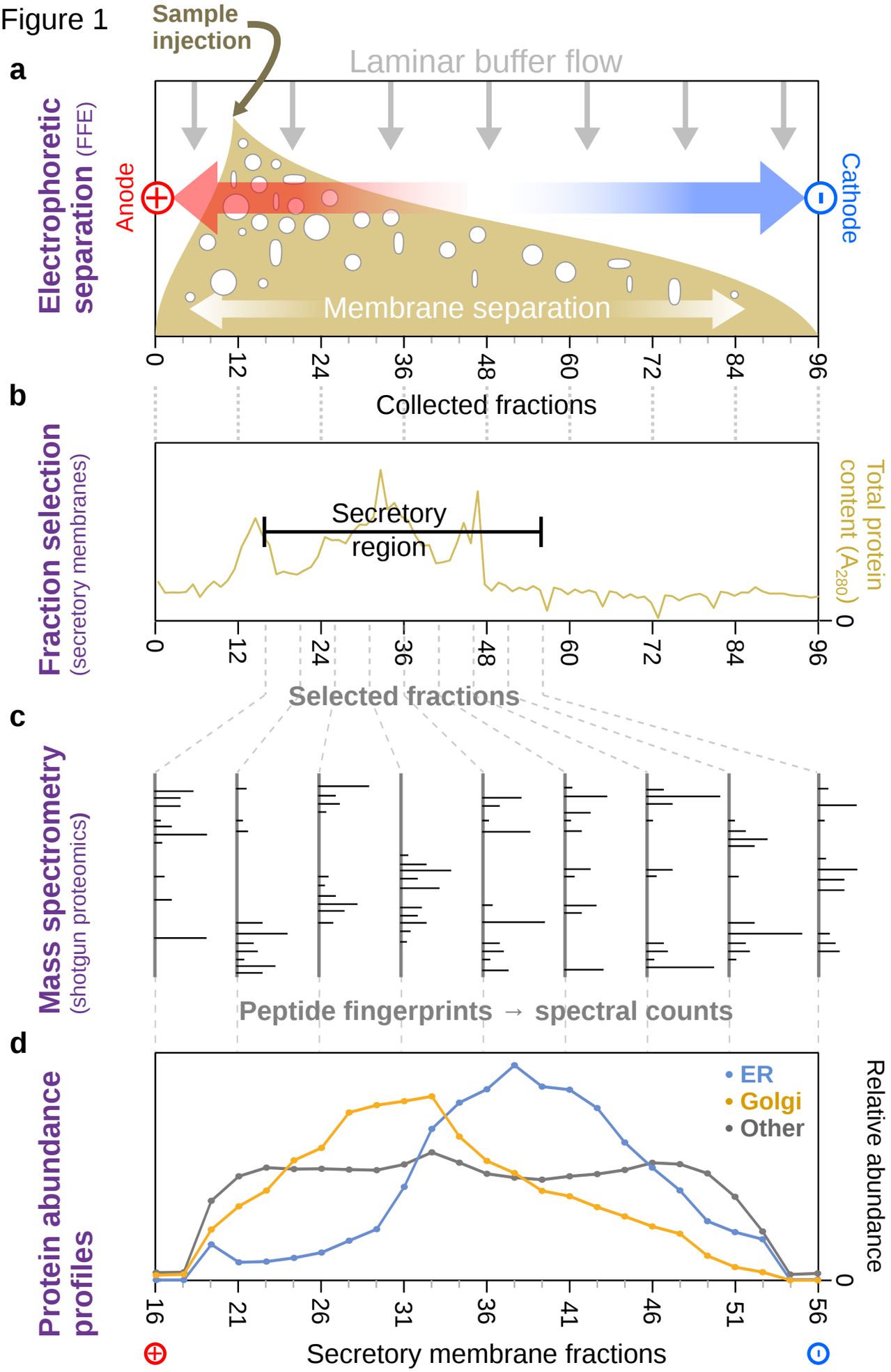


Figure 1. Schematic overview of electrophoretic separation profile analysis of endomembrane proteins.
a. Samples from Arabidopsis cell-suspension cultures, enriched in intact endomembranes, were separated by voltage under laminar flow, i.e. using free-flow electrophoresis (FFE). This provided gentle separation of membrane-bound compartments, according to their surface charges, and resulted in 96 separately collected fractions, ordered along the voltage axis.
b. Total protein content of FFE fractions was determined via absorption at 280 nm to identify the range of fractions with major endomembrane protein enrichment. These and adjacent fractions were then taken forward for more detailed analysis. Non-membrane components from the samples peaked in early fractions outside this range.
c. Endomembrane fractions were primarily investigated using shotgun proteomics, to measure the relative amounts of the different proteins contained therein. Here proteins were identified via the mass fingerprints of trypic digest peptides searched against the most recent Arabidopsis proteome using MASCOT software.
d. Average FFE abundance profiles for resident proteins from Golgi, ER and other organelles, using independent sub-cellular localizations derived from LOPIT analysis (see Table S1). Protein abundance values from multiple replicate FFE runs, in the form of MS spectral intensities, were combined (see Methods for the fraction matching procedure), generating 25 merged, consensus endomembrane fractions. Combined data is shown for totals of 200 ER, 204 Golgi and 1290 other organelle proteins.

Figure 2

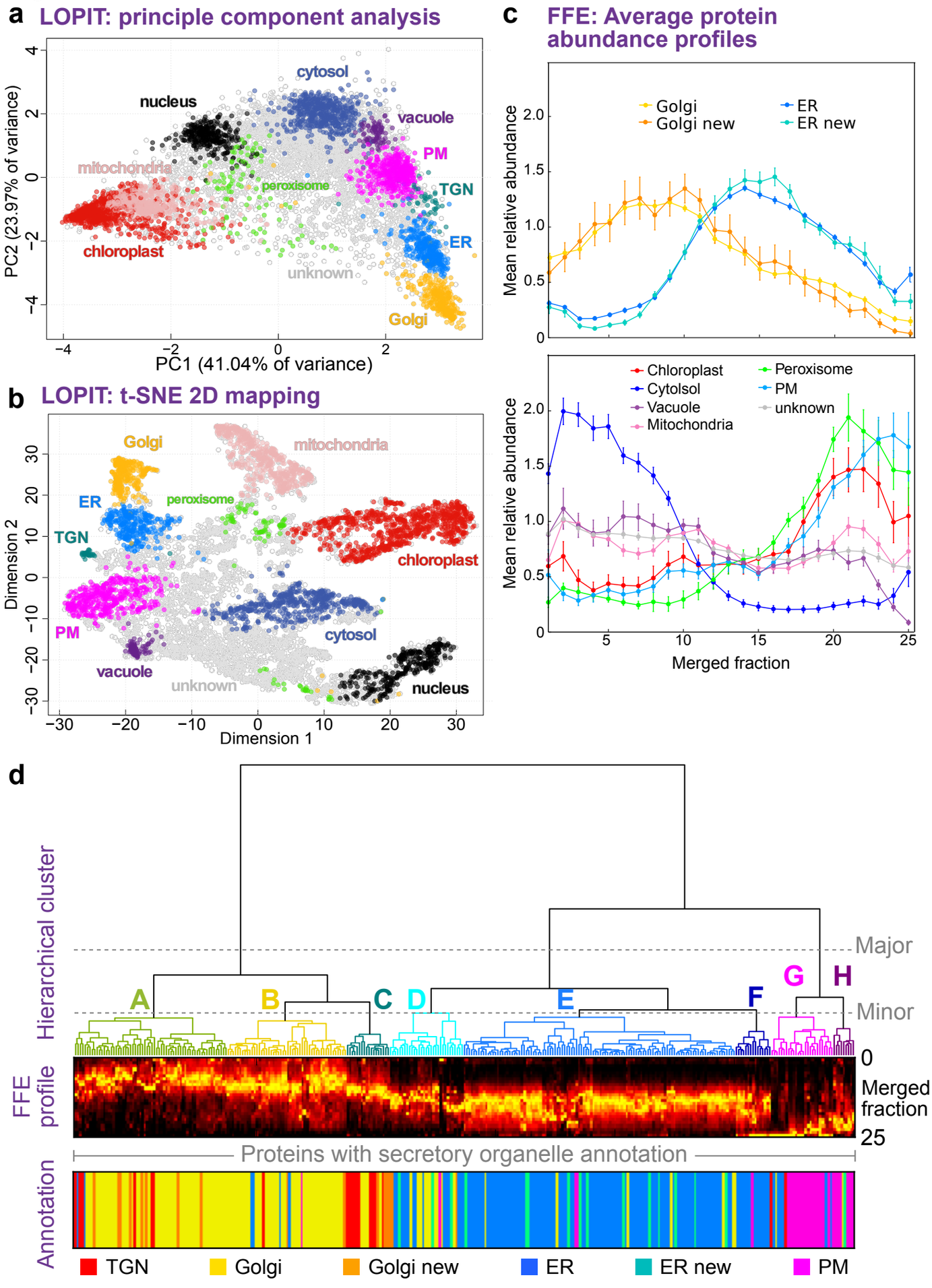


Figure 2. Primary determination of organelle sub-proteomes.

a. PCA analysis of a single LOPIT experiment: protein abundance profiles from density-based separation are presented by projection onto their two principle, orthogonal axes, representing most inter-protein variance. Each point represents a single protein, which is colored according to its organelle classification. Organelle clusters were distinguished using multi-class SVM on complete abundance profiles and used existing annotations for classification (see Methods).

b. Presentation of the same LOPIT data and classifications shown in a., presented as a two-dimensional t-SNE plot. This visualization attempts to preserve the proximity of similar profiles, and the separation of distinct profiles, over all data dimensions (whole profiles). This is unlike PCA which shows (dis)similarity along the selected projection axes.

c. Average FFE profiles, across 25 merged fractions from replicates R3-5 are shown for organelle groups classified using LOPIT data. Plotted values represent the mean abundance for each fraction in each organelle class, from per-protein normalized profiles (see Methods). Error bars represent the standard error of the mean. Data is shown separately for the ER and Golgi (upper plot), which peak as a class in central fractions, and the distinct profiles for other organelles/ compartments (lower plot). ER and Golgi proteomes have been sub-divided as either those belonging to the initial organelle markers or those newly classified as organelle residents, demonstrating the accuracy with which new residents were assigned.

d. Hierarchical clustering of secretory (ER, Golgi TGN and PM) protein FFE profiles. Merged abundance profiles from proteins identified in high-quality replicates R3-5 were clustered using Ward's method and presented as a dendrogram with the corresponding, underlying abundance profiles shown beneath as a color density plot, together with primary organelle classifications derived from LOPIT. The three major clusters that separated profiles generally into Golgi/TGN, ER and PM were further separated into eight smaller clusters, labelled A-H. Here a threshold was chosen so that each major ER and Golgi cluster contained three minor clusters.

Figure 3

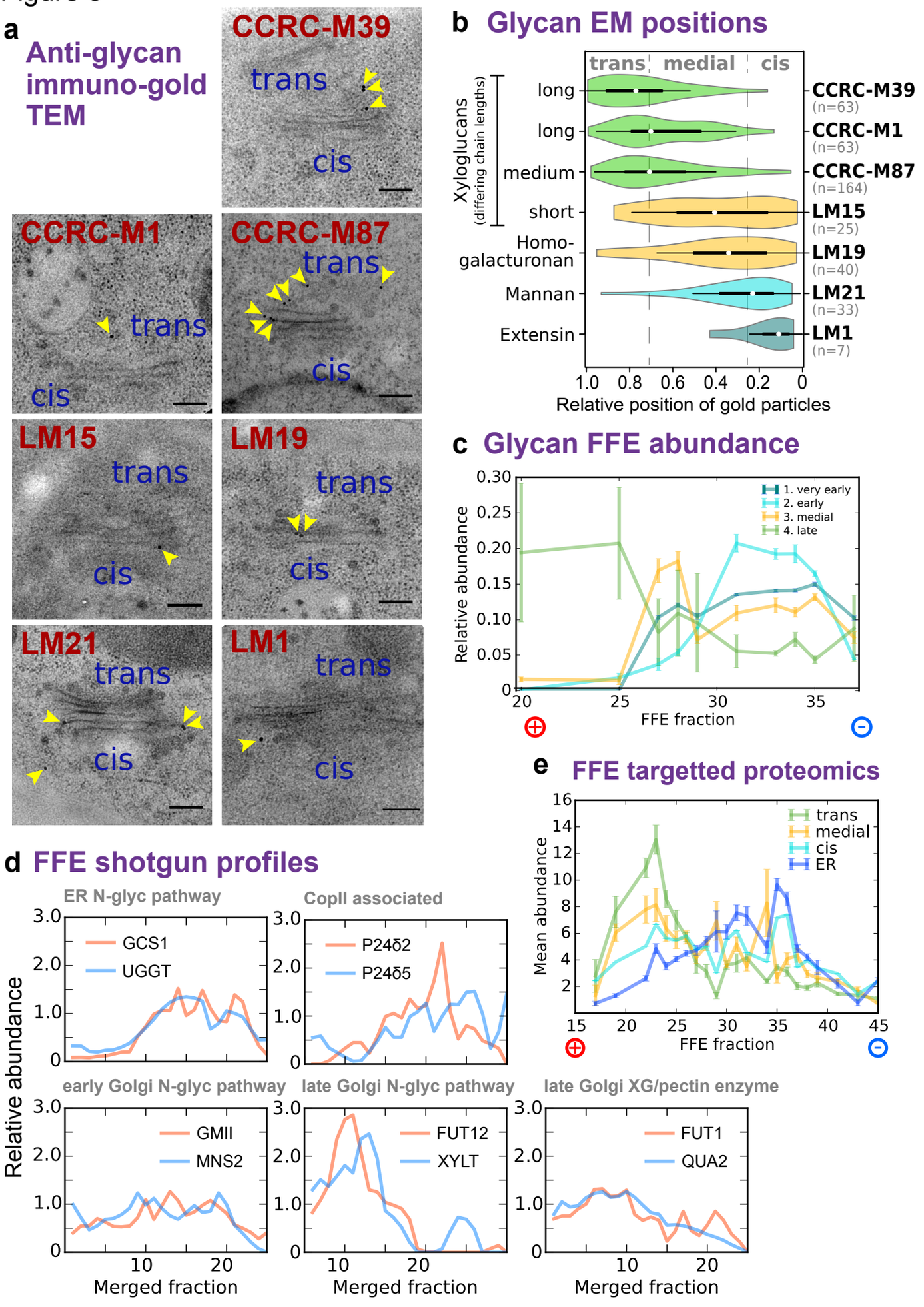
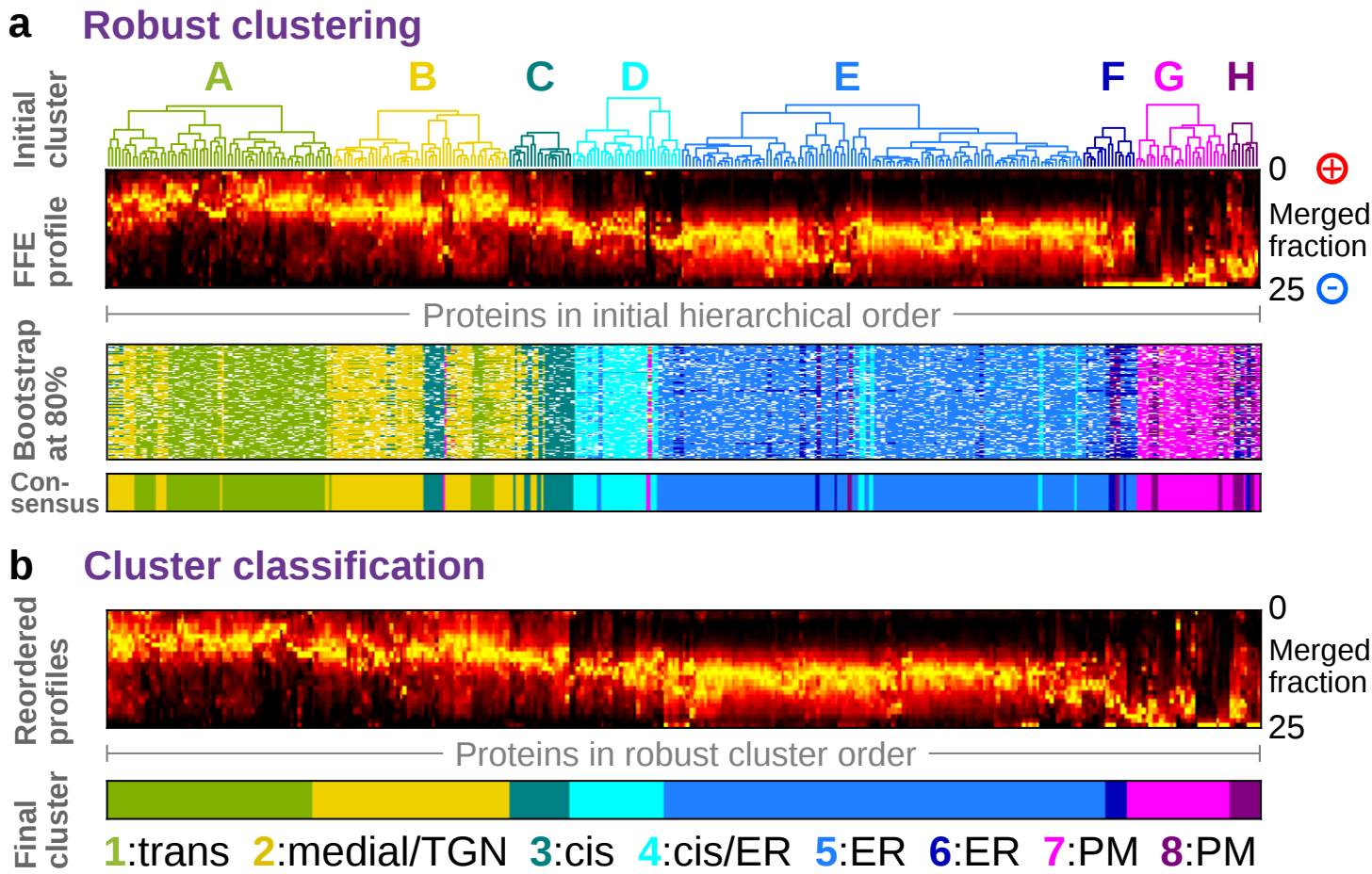
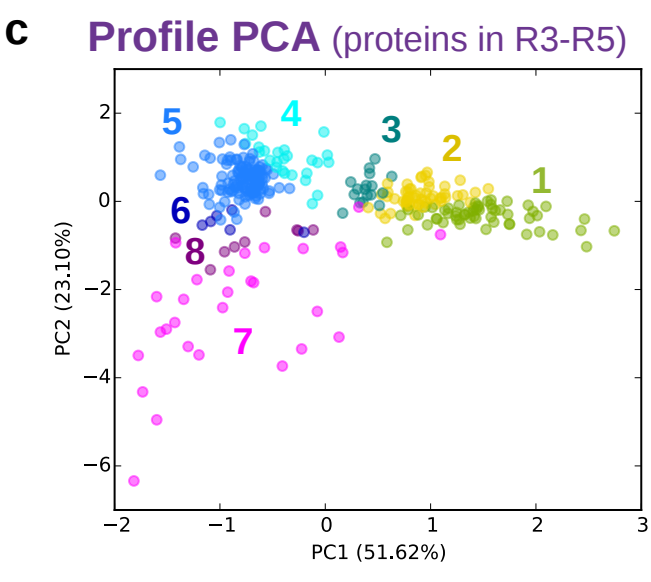
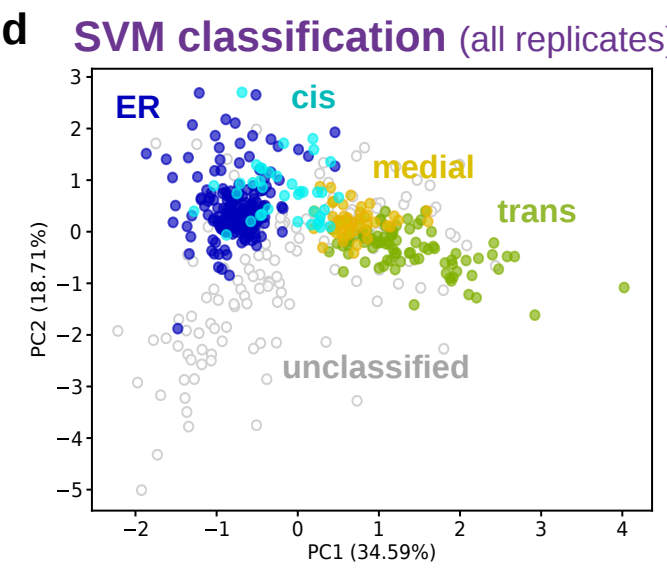


Figure 3. Establishing characteristics of early and late Golgi FFE profiles.
a. Example negative-stain TEM images showing the in-vivo distributions of several glycan epitopes, with varying structural complexity, across the Golgi stack. Glycans were localized using monoclonal antibodies liked to gold particles. All stacks are depicted with cis at the bottom and trans as the top, is indicated.
b. Violin plots showing the overall data from the immuno-gold TEM localization of glycan epitopes, as illustrated in a. The relative Golgi stack positions of gold particles represent the fraction of the particle distance to the outer cis face as a proportion of the total cis-trans thickness.
c. FFE abundance profiles for four classes of glycan epitope, with varying structural complexity. Class members and epitope structures are detailed in Table S2. Data is shown for detergent-extracted samples from FFE replicate R1 that were printed onto nitrocellulose microarrays and probed via alkaline phosphatase linked monoclonal antibodies. Error bars show SEM for n=3 antibodies (Group 4), n=9 (Group 3), n=2 (Group2) and n=5 (Group 1).
d. Exemplar FFE protein abundance profiles, as detected by high-throughput shotgun proteomics. Example proteins were selected on the basis of previously established sub-Golgi, ER and transitional ER-Golgi localization relating to well-known biomolecular functions in the secretory pathway.
e. FEE abundance profiles of selected proteins detected via high-sensitivity, targeted proteomics. Proteins (see Table S2) were chosen given an established function and localization specific to Golgi cisternae or the ER. Two independent peptides per protein were measured for n=7 (ER), n=1 (cis-Golgi), n=5 (medial Golgi) and n=3 (trans-Golgi) proteins. Solid lines indicate mean abundance over all proteins in the class and error bars represent SEM.

Figure 4







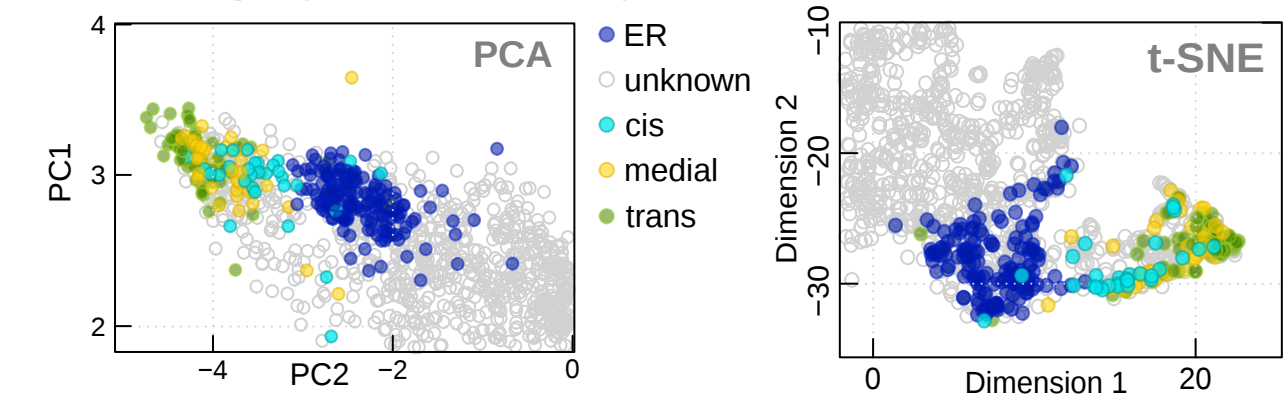


Figure 4. Classification of sub-Golgi compartments

a. Robust clustering of secretory protein FFE profiles via bootstrapping. Abundance profiles (second from top) were re-clustered using Ward's method 120 times, each time omitting 20% of the proteins. The resulting clusters were assigned to the corresponding initial clusters A-H (see Fig. 2) by similarity to the cluster medioids. These clusters are shown as a color map (third panel) where each row corresponds to a different, random subset of proteins, and is presented in the initial hierarchical cluster column order (as used in Fig. 2). The robust, consensus clusters (lower panel) were defined as the most common cluster identity for each protein over all the bootstrap trials.

b. FFE profiles for each of the eight consensus groups were separately re-clustered (Ward's method) to clearly visualize profile characteristics of each group. The groups were re-labelled 1-8 to discriminate them from the initial clusters A-H, which have (slightly) different memberships. These were then used for tentative assignment of particular groups (1-4) to sub-Golgi compartments using trends presented in Fig. 3. Abundance profiles are presented as a color density map, as in a, but in a new intra-group order.

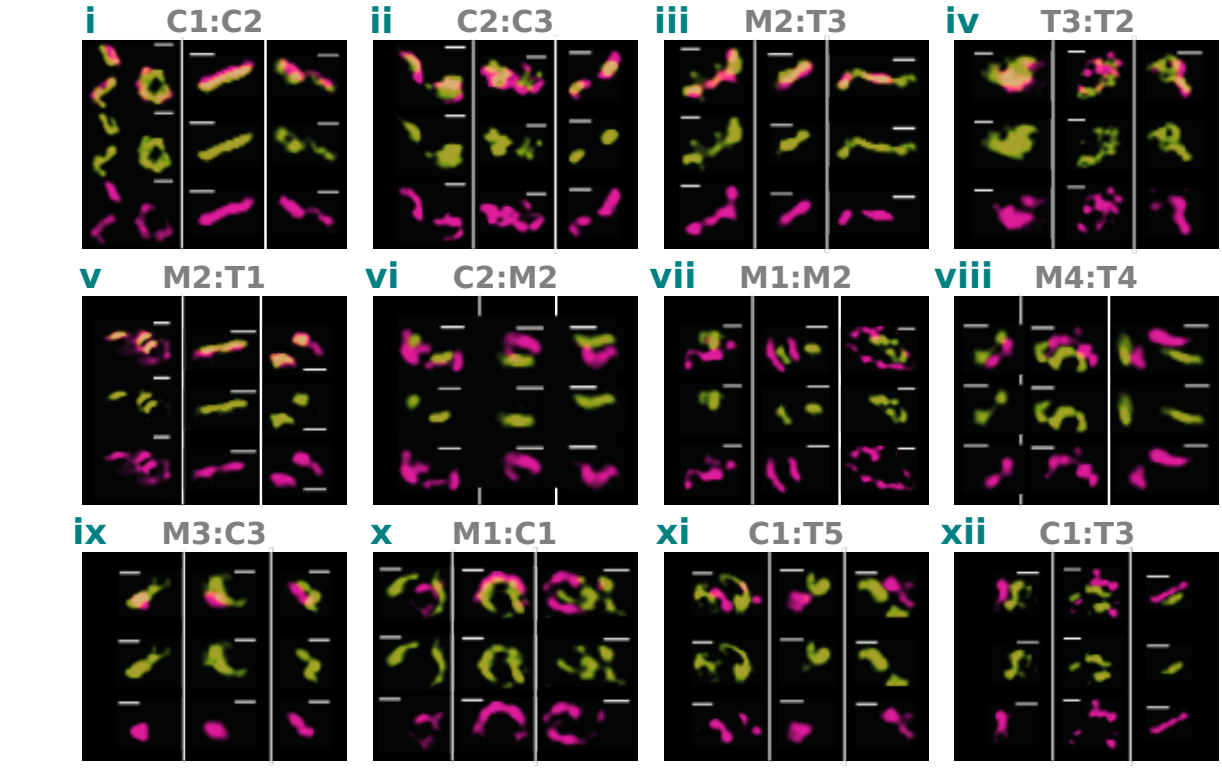
c. Merged FFE profile data, for proteins present in replicates R3-5, plotted as a two-dimensional PCA projection and labelled according to the bootstrap consensus clusters 1-8, as illustrated in b.

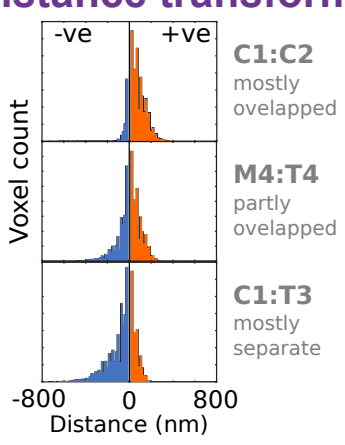
d. Merged FFE profile data for all secretory proteins detected in any of the replicates R1-5, presented as a two-dimensional PCA projection. Multi-class SVM was used to classify proteins (on whole FFE profiles, not the 2-D map) into three sub-Golgi groups and an ER group. The group labels used in the classification came from LOPIT to provide distinction between resident ER and Golgi proteins (and to exclude TGN ones), given that profiles overlap, to a degree, in the FFE data, but not in the LOPIT data. The consensus FFE sub-clusters (as in c) were then used to classify the three sub-Golgi groups from among the larger Golgi proteome. Consensus sub-clusters and final proteomes are detailed in Table S3.

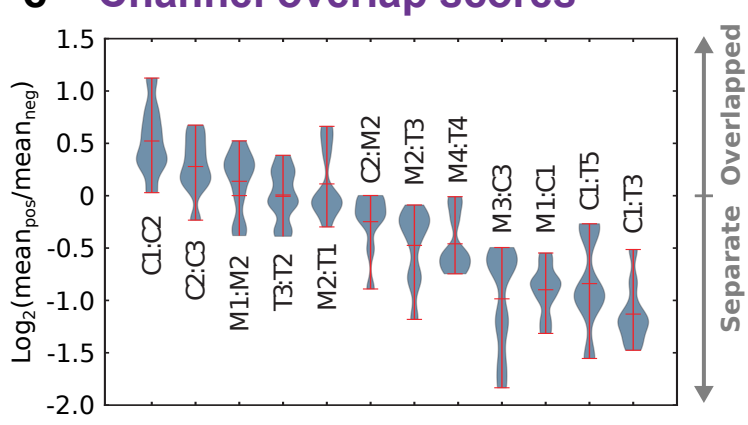
e. Re-presentation of a section of the LOPIT PCA map shown in Fig. 2a, now colored according to ER and sub-Golgi classes presented in d.

f. Re-presentation of a section of the 2D t-SNE map shown in Fig. 2b, now colored according to ER and sub-Golgi classes presented in d.

Figure 5







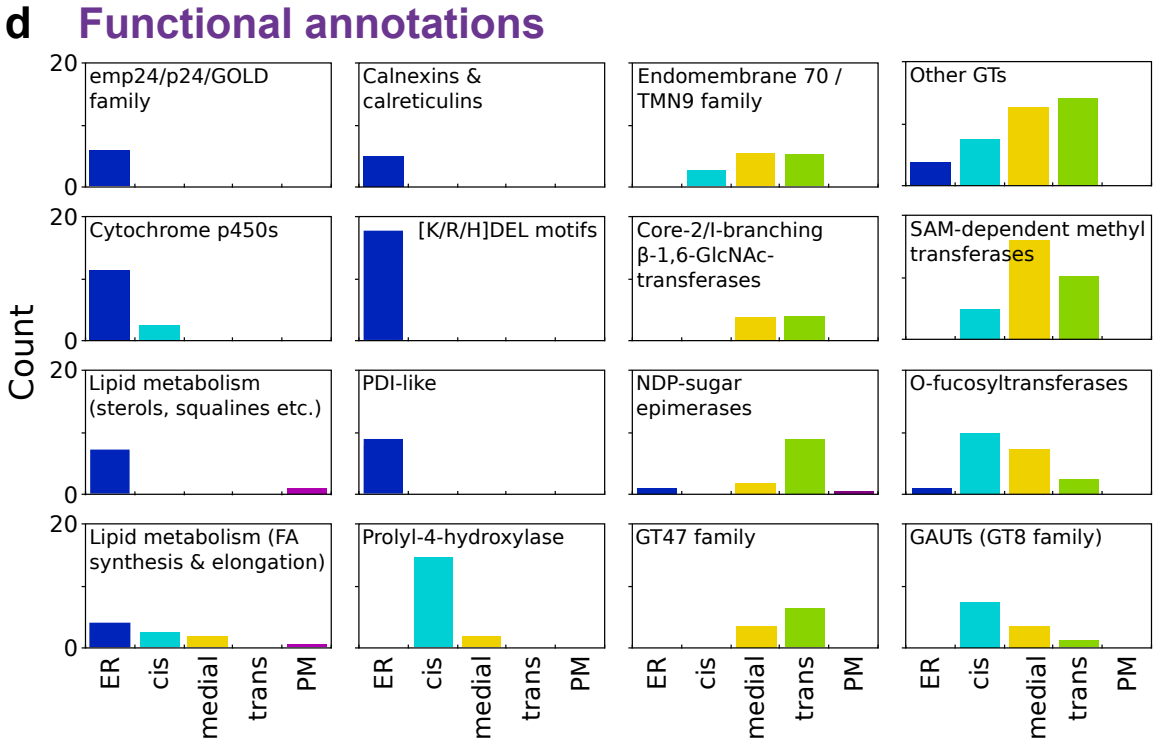


Figure 5. Validation of sub-Golgi protein localization

a. Example images of structured illumination microscopy (SIM) of validatory protein pairs, representative of cis (C), medial (M) and trans (T) Golgi sub-localizations. Sub-Golgi locations of PUBQ10 driven, C-terminally tagged GFP and RFP fusion proteins (88) were assayed to provide pairwise comparisons by using transient expression in N. bethamiana. For each protein pair, localization data were collected from nine regions (Dataset S5), incorporating three image stacks from at least two leaves per plant. Localizations were visualized in a single Golgi body from each of the three image stacks. The gene identifiers for the proteins were: AT2G20810.1 (C1), AT5G47780.1 (C2), AT2G43080.1 (C3), AT1G26850.1 (M1), AT3G62720.1 (M2), AT5G18480.1 (M3), AT1G19360.1 (M4), AT1G74380.1 (T1), AT1G08660.1 (T2), AT4G36890.1 (T4), AT2G35100.1 (T3), AT5G11730.1 (T5). Scale bars = 400 nm.

b. Three example histograms showing the distribution of distance transform values for image regions containing multiple Golgi stacks with spatially overlapping (top), partly overlapping (middle) and somewhat separate (bottom) labelled protein pairs, i.e. from red/green fluorescence microscopy illustrated in a. Channel signal overlap was quantified by thresholding intensities to generate regions of interest (ROIs), then summing the distance transform values for one channel's ROIs within the ROI bounds of the other. Here, negative values indicate greater separation and positive values indicate overlap.

c. The distribution of red/green channel overlap scores, over multiple image regions (n=9), for the validatory protein pairs shown in a, arranged in modal order. Overlap scores were calculated for each image region as the log2 ratio of mean absolute values either side of zero distance (see blue and orange regions in b), with positive values indicating more overlap. Image regions are given in Dataset S1.

d. Occurrence of proteins families and functional annotation in the secretory and sub-Golgi proteomes. Using ER, TGN and PM localizations derived from LOPIT data and sub-Golgi localizations from FFE (see Table S4), proteins were grouped variously according to family, MapMan (99) functional categorization and possession of the K/H/RDEL ER-retrieval motif. Groups with at least 5 members are presented here.

Figure 6

Figure 6: Comparison of Type II TM protein paralogues with different sub-Golgi classification.

Alignments are shown for pairs of similar, homologous proteins from Arabidopsis which have different sub-Golgi localisations. TM-span regions are idicated in bold. The blue Arg/Lys at the cytoplasmic edge highlight the start of the TM span. Phe residues are coloured either pink or cyan to indicate relative position in the TM span. Within 15 residues of the exoplasmic TM edge Ser residues are coloured yellow and three consecutive Ser are red.

Figure 7

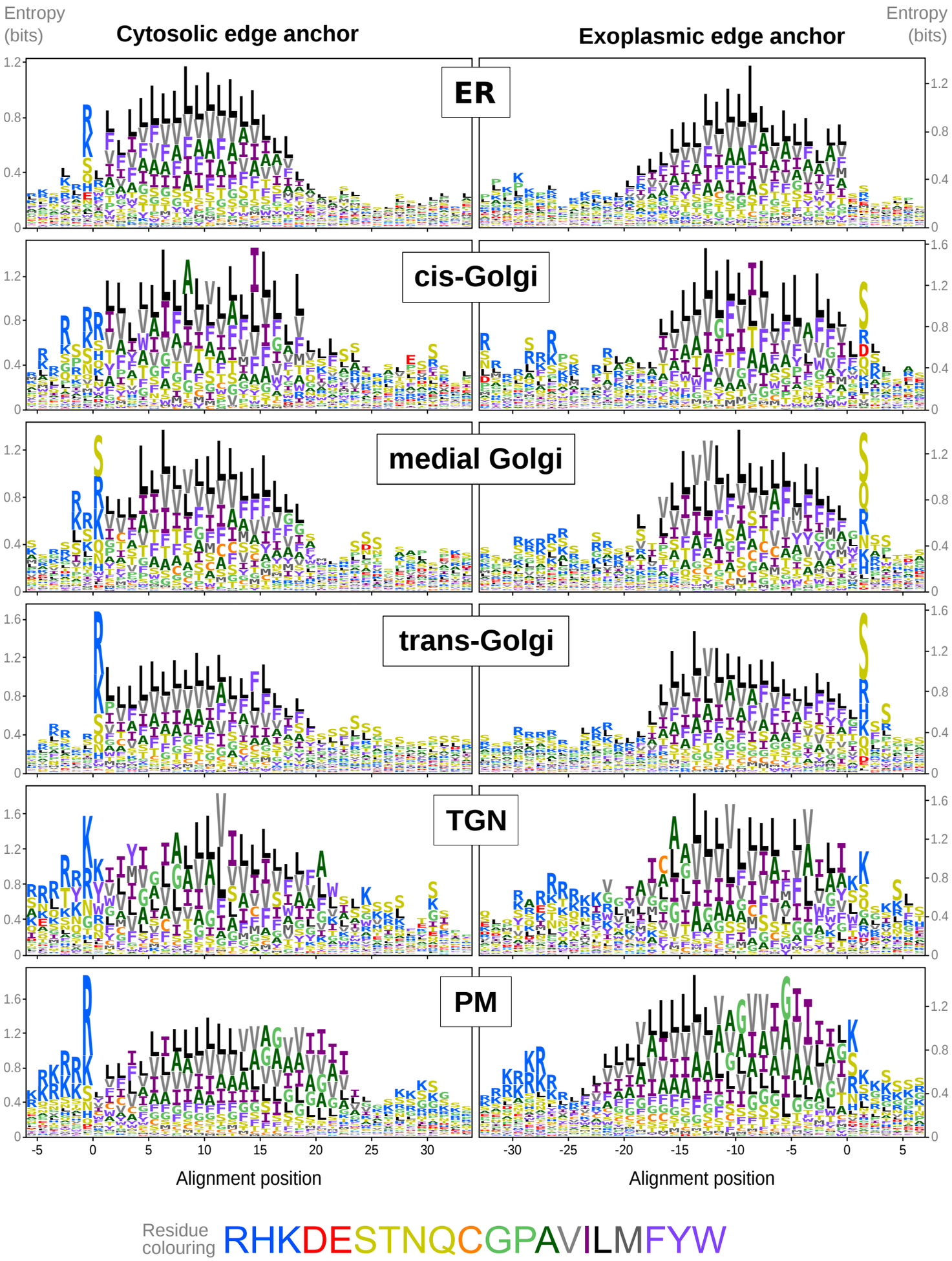


Figure 7. Transmembrane amino acid composition in sub-Golgi and secretory compartments.
Logo plots of single-span TM proteins from secretory and sub-Golgi proteomes indicating the relative abundance of amino acids at and around aligned TM spans.
Data is shown for the Arabidopsis proteins localized by LOPIT and FFE and their very close homologues. Different sequences were aligned at either the cytoplasmic (left column) or exoplasmic/luminal (right column) edge of the hydrophobic TM spans. (See Methods for details of gathering homologues and aligning TM sequences). The different amino acids are color coded according to their physiochemical properties, as indicated in the color key (bottom). Logo plots were generated after randomly sampling 1000 sequences for each dataset, from position-specific residue abundance probabilities calculated from dissimilarity weighted sequences. This was done to reduce the bias caused by the different sizes of protein families, i.e. which are informatically somewhat redundant.

Figure 8

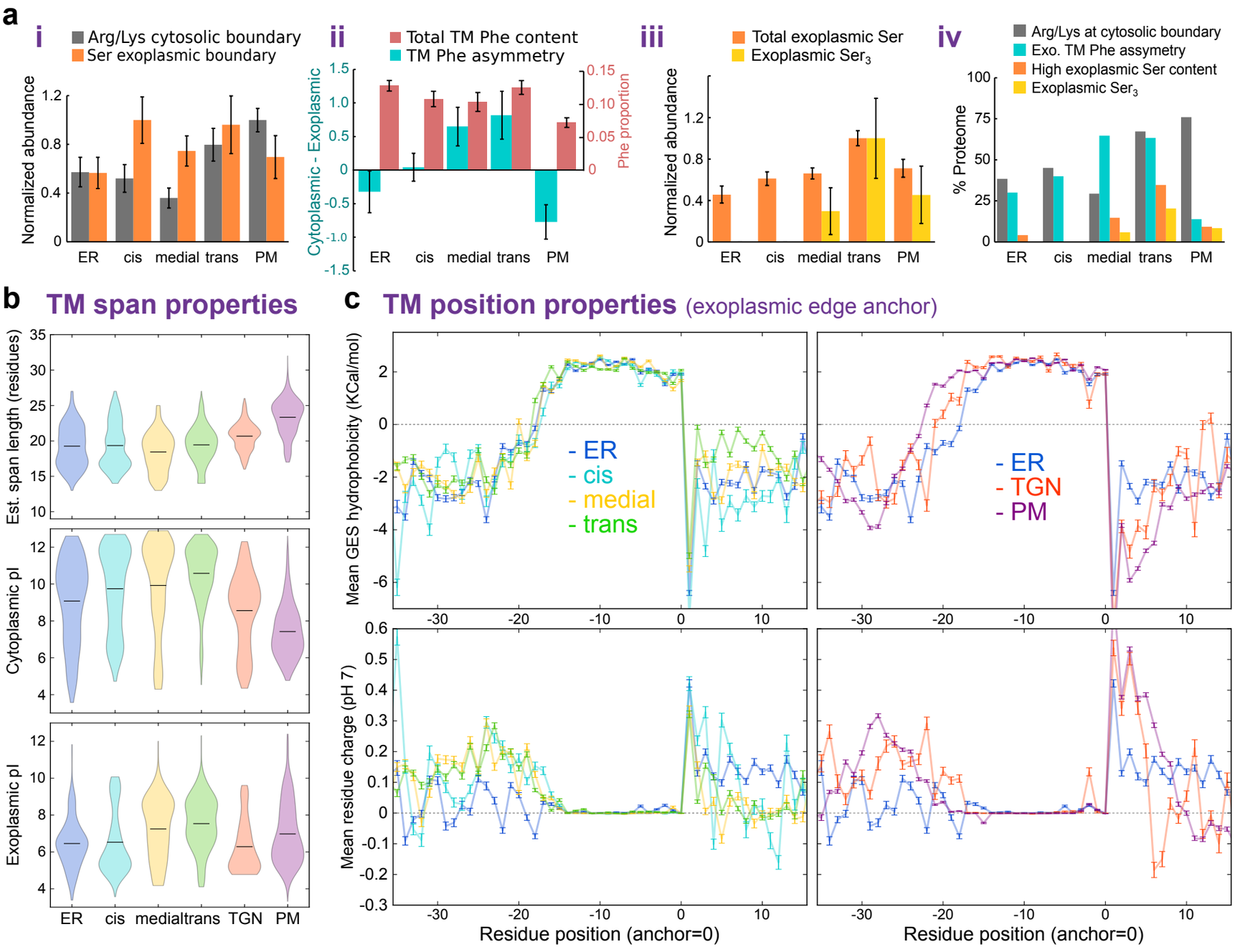


Figure 8. Comparison of protein sequence features in organelle and sub-organelle proteomes.
a. Abundance of sequence features at and around the TM spans of single-span proteins in the secretory and sub-Golgi proteomes. Data is shown for 63 ER, 23 cis-, 37 medial and 54 trans-Golgi proteins, and 108 PM transmembrane proteins. i. The relative abundance of lysine or arginine at the cytoplasmic TM edge and serine at the exoplasmic/luminal edge. Values were normalized relative to the maximum observation. ii. Overall TM phenylalanine content, as a proportion of TM span length, and the cytoplasmic-exoplasmic asymmetry of TM phenylaniline; asymmetry was calculated as the difference in the abundance between the two halves of each TM span. iii. The relative abundance of Serine and presence of three or more consecutive Serines in the 15-residue exoplasmic region immediately flanking the TM spans. Values were normalized relative to the maximum observation. iv. An overview of the results presented in i-iii, but shown as a proportion of each sub-cellular proteome. Here phenylalanine asymmetry corresponded to positive values presented in ii and high serine content corresponded to a count of at least 5 in the 15 flanking exoplasmic residues. For panels i-iii bar heights are mean values and errors represent the standard error in the mean.
b. Distributions TM span properties for different sub-proteome groups. Datasets for localized single-span TM proteins from Arabidopsis were expanded through close homology searches (as used in Figure 6), where sequence contributions were weighted by dissimilarity and TM-span were edges defined, as detailed in the Methods. TM-span length (top), pI of the entire cytoplasmic region (middle) and pI of the entire exoplasmic region (bottom) are shown as violin plots for different secretory and sub-Golgi compartments (defined by LOPIT and FFE respectively).
c. Line plots of per-position TM hydrophobicity (top) and mean residue charge (bottom) for localized Arabidopsis and homologue over TM hydrophobic core and flanking regions (as in Figure 6). TM spans were anchored at their exoplasmic boundary. Plotted values represent the means at each TM aligned position, over different, dissimilarity-weighted proteins. Error bars represent the standard error in the mean.

Parsed Citations

1. Klute MJ, Melançon P, Dacks JB. Evolution and diversity of the Golgi. Cold Spring Harb Perspect Biol. 2011 Aug 1;3(8):a007849.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

2. Strasser R. Plant protein glycosylation. Glycobiology. 2016 Sep;26(9):926–939.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

3. Van de Meene AML, Doblin MS, Bacic A. The plant secretory pathway seen through the lens of the cell wall. Protoplasma. 2017 Jan;254(1):75–94.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

4. Ito Y, Uemura T, Nakano A. Formation and maintenance of the Golgi apparatus in plant cells. Int Rev Cell Mol Biol. 2014;310:221–287.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

5. Brandizzi F, Barlowe C. Organization of the ER-Golgi interface for membrane traffic control. Nat Rev Mol Cell Biol. 2013 Jun;14(6):382–392.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

6. Gendre D, Jonsson K, Boutté Y, Bhalerao RP. Journey to the cell surface--the central role of the trans-Golgi network in plants. Protoplasma. 2015 Mar;252(2):385–398.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

7. Robinson DG, Pimpl P. Clathrin and post-Golgi trafficking: a very complicated issue. Trends Plant Sci. 2014 Mar;19(3):134–139.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

8. Xiang L, Etxeberria E, Van den Ende W. Vacuolar protein sorting mechanisms in plants. FEBS J. 2013 Feb;280(4):979–993.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

9. Heard W, Sklenář J, Tomé DFA, Robatzek S, Jones AME. Identification of Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection. Mol Cell Proteomics. 2015 Jul;14(7):1796–1813.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

10. Hawes C, Kiviniemi P, Kriechbaumer V. The endoplasmic reticulum: a dynamic and well-connected organelle. J Integr Plant Biol. 2015 Jan;57(1):50–62.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

11. Sharpe HJ, Stevens TJ, Munro S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell. 2010 Jul 9;142(1):158–169.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

12. Banfield DK. Mechanisms of protein retention in the Golgi. Cold Spring Harb Perspect Biol. 2011 Aug 1;3(8):a005264.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

13. Woo CH, Gao C, Yu P, Tu L, Meng Z, Banfield DK, et al. Conserved function of the lysine-based KXD/E motif in Golgi retention for endomembrane proteins among different organisms. Mol Biol Cell. 2015 Nov 15;26(23):4280–4293.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

14. Gao C, Cai Y, Wang Y, Kang B-H, Aniento F, Robinson DG, et al. Retention mechanisms for ER and Golgi membrane proteins. Trends Plant Sci. 2014 Aug;19(8):508–515.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

15. Saint-Jore-Dupas C, Nebenführ A, Boulaflous A, Follet-Gueye M-L, Plasson C, Hawes C, et al. Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. Plant Cell. 2006 Nov 30;18(11):3182–3200.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

16. Schoberer J, Strasser R. Sub-compartmental organization of Golgi-resident N-glycan processing enzymes in plants. Mol Plant. 2011 Mar;4(2):220–228.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

17. Tie HC, Mahajan D, Chen B, Cheng L, VanDongen AMJ, Lu L. A novel imaging method for quantitative Golgi localization reveals differential intra-Golgi trafficking of secretory cargoes. Mol Biol Cell. 2016 Mar 1;27(5):848–861.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

18. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlén M, et al. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods. 2013 Apr;10(4):315–323.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

19. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, et al. Quantitative proteomics analysis of the secretory pathway. Cell. 2006 Dec 15;127(6):1265–1281.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

20. Mulvey CM, Breckels LM, Geladaki A, Britovšek NK, Nightingale DJH, Christoforou A, et al. Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. Nat Protoc. 2017 Jun;12(6):1110–1135.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

21. Groen AJ, Sancho-Andrés G, Breckels LM, Gatto L, Aniento F, Lilley KS. Identification of trans-golgi network proteins in Arabidopsis thaliana root tissue. J Proteome Res. 2014 Feb 7;13(2):763–776.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

22. Dunkley TPJ, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, et al. Mapping the Arabidopsis organelle proteome. Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518–6523.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

23. Nikolovski N, Rubtsov D, Segura MP, Miles GP, Stevens TJ, Dunkley TPJ, et al. Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. Plant Physiol. 2012 Oct;160(2):1037–1051.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

24. Islinger M, Eckerskorn C, Völkl A. Free-flow electrophoresis in the proteomic era: a technique in flux. Electrophoresis. 2010 Jun;31(11):1754–1763.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

25. Barkla BJ, Vera-Estrella R, Pantoja O. Enhanced separation of membranes during free flow zonal electrophoresis in plants. Anal Chem. 2007 Jul 15;79(14):5181–5187.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

26. Morré DJ, Mollenhauer HH, editors. Isolation and Subfractionation. The Golgi Apparatus. New York, NY: Springer New York; 2009. p. 39–61.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

27. Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, et al. Isolation and proteomic characterization of the Arabidopsis Golgi defines functional and novel components involved in plant cell wall biosynthesis. Plant Physiol. 2012 May;159(1):12–26.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

28. Breckels LM, Mulvey CM, Lilley KS, Gatto L. A Bioconductor workflow for processing and analysing spatial proteomics data [version 1; peer review: 1 approved, 1 approved with reservations]. F1000Res. 2016 Dec 28;5.

29. Christoforou A, Mulvey CM, Breckels LM, Geladaki A, Hurrell T, Hayward PC, et al. A draft map of the mouse pluripotent stem cell spatial proteome. Nat Commun. 2016 Jan 12;7:8992.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

30. Gatto L, Breckels LM, Wieczorek S, Burger T, Lilley KS. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. Bioinformatics. 2014 May 1;30(9):1322–1324.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

31. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science. 2017 May 26;356(6340).

32. Breckels LM, Gatto L, Christoforou A, Groen AJ, Lilley KS, Trotter MWB. The effect of organelle discovery upon sub-cellular

protein localisation. J Proteomics. 2013 Aug 2;88:129–140.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

33. Breckels LM, Holden SB, Wojnar D, Mulvey CM, Christoforou A, Groen A, et al. Learning from Heterogeneous Data Sources: An Application in Spatial Proteomics. PLoS Comput Biol. 2016 May 13;12(5):e1004920.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

34. Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. Nucleic Acids Res. 2017 Jan 4;45(D1):D1064–D1074.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

35. Van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res [Internet]. 2008 Nov;9:2579–2605. Available from: http://www.jmlr.org/papers/v9/vandermaaten08a.html
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

36. Hooper CM, Stevens TJ, Saukkonen A, Castleden IR, Singh P, Mann GW, et al. Multiple marker abundance profiling: combining selected reaction monitoring and data-dependent acquisition for rapid estimation of organelle abundance in subcellular samples. Plant J. 2017 Dec;92(6):1202–1217.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

37. Eubel H, Lee CP, Kuo J, Meyer EH, Taylor NL, Millar AH. Free-flow electrophoresis for purification of plant mitochondria by surface charge. Plant J. 2007 Nov;52(3):583–594.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

38. De Michele R, McFarlane HE, Parsons HT, Meents MJ, Lao J, González Fernández-Niño SM, et al. Free-Flow Electrophoresis of Plasma Membrane Vesicles Enriched by Two-Phase Partitioning Enhances the Quality of the Proteome from Arabidopsis Seedlings. J Proteome Res. 2016 Mar 4;15(3):900–913.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

39. Zielinska DF, Gnad F, Schropp K, Wiśniewski JR, Mann M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. Mol Cell. 2012 May 25;46(4):542–548.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

40. Yeats TH, Bacic A, Johnson KL. Plant glycosylphosphatidylinositol (GPI) anchored proteins at the plasma membrane-cell wall nexus. J Integr Plant Biol. 2018 Apr 18;60(8):649–669.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

41. Driouich A, Follet-Gueye M-L, Bernard S, Kousar S, Chevalier L, Vicré-Gibouin M, et al. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Front Plant Sci. 2012 Apr 30;3:79.

42. Viotti C, Bubeck J, Stierhof Y-D, Krebs M, Langhans M, van den Berg W, et al. Endocytic and secretory traffic in Arabidopsis merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle. Plant Cell. 2010 Apr 30;22(4):1344–1357.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

43. Marcus SE, Verhertbruggen Y, Hervé C, Ordaz-Ortiz JJ, Farkas V, Pedersen HL, et al. Pectic homogalacturonan masks abundant sets of xyloglucan epitopes in plant cell walls. BMC Plant Biol. 2008 May 22;8:60.

44. Smallwood M, Beven A, Donovan N, Neill SJ, Peart J, Roberts K, et al. Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex. Plant J. 1994 Feb;5(2):237–246.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

45. Okekeogbu IO, Pattathil S, González Fernández-Niño SM, Aryal UK, Penning BW, Lao J, et al. Glycome and proteome components of Golgi membranes are common between two angiosperms with distinct cell wall structures. Plant Cell. 2019 Mar 26;

46. Wilkop T, Pattathil S, Ren G, Davis DJ, Bao W, Duan D, et al. A Hybrid Approach Enabling Large-Scale Glycomic Analysis of Post-Golgi Vesicles Reveals a Transport Route for Polysaccharides. Plant Cell. 2019 Mar;31(3):627–644.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

47. Ralet M-C, Tranquet O, Poulain D, Moïse A, Guillon F. Monoclonal antibodies to rhamnogalacturonan I backbone. Planta. 2010 May;231(6):1373–1383.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

48. Nilsson T, Au CE, Bergeron JJM. Sorting out glycosylation enzymes in the Golgi apparatus. FEBS Lett. 2009 Dec 3;583(23):3764–3769.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

49. Gao P, Xin Z, Zheng Z-L. The OSU1/QUA2/TSD2-encoded putative methyltransferase is a critical modulator of carbon and nitrogen nutrient balance response in Arabidopsis. PLoS ONE. 2008 Jan 2;3(1):e1387.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

50. Chevalier L, Bernard S, Ramdani Y, Lamour R, Bardor M, Lerouge P, et al. Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. Plant J. 2010 Dec;64(6):977–989.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

51. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell. 2009 Aug 21;138(4):795–806.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

52. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat Methods. 2010 Jan;7(1):43–46.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

53. Heintzmann R, Huser T. Super-Resolution Structured Illumination Microscopy. Chem Rev. 2017 Dec 13;117(23):13890–13908.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

54. Yuasa K, Toyooka K, Fukuda H, Matsuoka K. Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. Plant J. 2005 Jan;41(1):81–94.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

55. Xu H, Su W, Cai M, Jiang J, Zeng X, Wang H. The asymmetrical structure of Golgi apparatus membranes revealed by in situ atomic force microscope. PLoS ONE. 2013 Apr 16;8(4):e61596.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

56. Pattathil S, Avci U, Baldwin D, Swennes AG, McGill JA, Popper Z, et al. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. Plant Physiol. 2010 Jun;153(2):514–525.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

57. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, et al. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc Natl Acad Sci U S A. 2011 Dec 13;108(50):20225–20230.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

58. Iwai H, Hokura A, Oishi M, Chida H, Ishii T, Sakai S, et al. The gene responsible for borate cross-linking of pectin Rhamnogalacturonan-II is required for plant reproductive tissue development and fertilization. Proc Natl Acad Sci U S A. 2006 Oct 31;103(44):16592–16597.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

59. Jensen JK, Sørensen SO, Harholt J, Geshi N, Sakuragi Y, Møller I, et al. Identification of a xylogalacturonan xylosyltransferase involved in pectin biosynthesis in Arabidopsis. Plant Cell. 2008 May 6;20(5):1289–1302.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

60. Zhong R, Peña MJ, Zhou G-K, Nairn CJ, Wood-Jones A, Richardson EA, et al. Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell. 2005 Dec;17(12):3390–3408.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

61. Harholt J, Jensen JK, Verhertbruggen Y, Søgaard C, Bernard S, Nafisi M, et al. ARAD proteins associated with pectic Arabinan biosynthesis form complexes when transiently overexpressed in planta. Planta. 2012 Jul;236(1):115–128.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

62. Knoch E, Dilokpimol A, Tryfona T, Poulsen CP, Xiong G, Harholt J, et al. A β-glucuronosyltransferase from Arabidopsis thaliana involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. Plant J. 2013 Dec;76(6):1016–1029.

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

63. Velasquez SM, Ricardi MM, Dorosz JG, Fernandez PV, Nadra AD, Pol-Fachin L, et al. O-glycosylated cell wall proteins are essential in root hair growth. Science. 2011 Jun 17;332(6036):1401–1403.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

64. Chen Y, Dong W, Tan L, Held MA, Kieliszewski MJ. Arabinosylation Plays a Crucial Role in Extensin Cross-linking In Vitro. Biochemistry insights. 2015 Sep 20;8(Supple 2):1–13.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

65. Miao Y, Li H-Y, Shen J, Wang J, Jiang L. QUASIMODO 3 (QUA3) is a putative homogalacturonan methyltransferase regulating cell wall biosynthesis in Arabidopsis suspension-cultured cells. J Exp Bot. 2011 Oct;62(14):5063–5078.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

66. Krupková E, Immerzeel P, Pauly M, Schmülling T. The TUMOROUS SHOOT DEVELOPMENT2 gene of Arabidopsis encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J. 2007 May;50(4):735–750.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

67. Luini A. A brief history of the cisternal progression-maturation model. Cell Logist. 2011 Jan;1(1):6–11.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

68. Glick BS, Luini A. Models for Golgi traffic: a critical assessment. Cold Spring Harb Perspect Biol. 2011 Nov 1;3(11):a005215.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

69. Donohoe BS, Kang B-H, Gerl MJ, Gergely ZR, McMichael CM, Bednarek SY, et al. Cis-Golgi cisternal assembly and biosynthetic activation occur sequentially in plants and algae. Traffic. 2013 May;14(5):551–567.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

70. Schoberer J, Liebminger E, Vavra U, Veit C, Grünwald-Gruber C, Altmann F, et al. Golgi localization of GnTI requires a polar amino acid residue within its transmembrane domain. Plant Physiol. 2019 Apr 10;

71. Leventis PA, Grinstein S. The distribution and function of phosphatidylserine in cellular membranes. Annu Rev Biophys. 2010;39:407–427.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

72. Simon MLA, Platre MP, Assil S, van Wijk R, Chen WY, Chory J, et al. A multi-colour/multi-affinity marker set to visualize phosphoinositide dynamics in Arabidopsis. Plant J. 2014 Jan;77(2):322–337.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

73. Poulsen LR, López-Marqués RL, McDowell SC, Okkeri J, Licht D, Schulz A, et al. The Arabidopsis P4-ATPase ALA3 localizes to the golgi and requires a beta-subunit to function in lipid translocation and secretory vesicle formation. Plant Cell. 2008 Mar 14;20(3):658–676.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

74. Simon MLA, Platre MP, Marquès-Bueno MM, Armengot L, Stanislas T, Bayle V, et al. A PtdIns(4)P-driven electrostatic field controls cell membrane identity and signalling in plants. Nature Plants. 2016 Jun 20;2:16089.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

75. Bogdanov M, Dowhan W, Vitrac H. Lipids and topological rules governing membrane protein assembly. Biochim Biophys Acta. 2014 Aug;1843(8):1475–1488.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

76. Martinière A, Bassil E, Jublanc E, Alcon C, Reguera M, Sentenac H, et al. In vivo intracellular pH measurements in tobacco and Arabidopsis reveal an unexpected pH gradient in the endomembrane system. Plant Cell. 2013 Oct 8;25(10):4028–4043.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

77. Quiroga R, Trenchi A, González Montoro A, Valdez Taubas J, Maccioni HJF. Short transmembrane domains with high-volume exoplasmic halves determine retention of Type II membrane proteins in the Golgi complex. J Cell Sci. 2013 Dec 1;126(Pt 23):5344–5349.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

78. Guo Y, Sirkis DW, Schekman R. Protein sorting at the trans-Golgi network. Annu Rev Cell Dev Biol. 2014 Aug 18;30:169–206.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

79. Surma MA, Klose C, Simons K. Lipid-dependent protein sorting at the trans-Golgi network. Biochim Biophys Acta. 2012 Aug;1821(8):1059–1067.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

80. Wattelet-Boyer V, Brocard L, Jonsson K, Esnay N, Joubès J, Domergue F, et al. Enrichment of hydroxylated C24- and C26-acyl-chain sphingolipids mediates PIN2 apical sorting at trans-Golgi network subdomains. Nat Commun. 2016 Sep 29;7:12788.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

81. Teese MG, Langosch D. Role of gxxxg motifs in transmembrane domain interactions. Biochemistry. 2015 Aug 25;54(33):5125–5135.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

82. Eubel H, Meyer EH, Taylor NL, Bussell JD, O'Toole N, Heazlewood JL, et al. Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of Arabidopsis cell culture peroxisomes. Plant Physiol. 2008 Dec;148(4):1809–1829.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

83. Griss J, Perez-Riverol Y, Lewis S, Tabb DL, Dianes JA, Del-Toro N, et al. Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. Nat Methods. 2016 Aug;13(8):651–656.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

84. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015 Aug 4;53(8):474–485.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

85. Team R. R: A language and environment for statistical computing. 2013; Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.5851&rep=rep1&type=pdf

86. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004 Sep 15;5(10):R80.

87. Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics. 2012 Jan 15;28(2):288–289.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

88. McFarlane HE, Young RE, Wasteneys GO, Samuels AL. Cortical microtubules mark the mucilage secretion domain of the plasma membrane in Arabidopsis seed coat cells. Planta. 2008 May;227(6):1363–1375.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

89. Staehelin LA, Giddings TH, Kiss JZ, Sack FD. Macromolecular differentiation of Golgi stacks in root tips of Arabidopsis and Nicotiana seedlings as visualized in high pressure frozen and freeze-substituted samples. Protoplasma. 1990;157(1-3):75–91.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

90. Pedersen HL, Fangel JU, McCleary B, Ruzanski C, Rydahl MG, Ralet M-C, et al. Versatile high resolution oligosaccharide microarrays for plant glycobiology and cell wall research. J Biol Chem. 2012 Nov 16;287(47):39429–39438.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

91. Batth TS, Singh P, Ramakrishnan VR, Sousa MML, Chan LJG, Tran HM, et al. A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. Metab Eng. 2014 Nov;26:48–56.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

92. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010 Apr 1;26(7):966–968.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

93. Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K, Blatt MR. A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J. 2010 Oct;64(2):355–365.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

94. Gustafsson MGL, Shao L, Carlton PM, Wang CJR, Golubovskaya IN, Cande WZ, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008 Jun;94(12):4957–4970.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

95. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jun 28;9(7):676–682.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

96. Kapur JN, Sahoo PK, Wong AKC. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing. 1985 Mar;29(3):273–285.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

97. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan;43(Database issue):D204–12.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

98. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011 Sep 29;8(10):785–786.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

99. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567–580.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

100. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004 May 14;338(5):1027–1036.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

101. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

102. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol. 2014;1079:105–116.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

103. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 2016 Jan 4;44(D1):D447–56.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

104. Ramsak Ž, Baebler Š, Rotter A, Korbar M, Mozetic I, Usadel B, et al. GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology. Nucleic Acids Res. 2014 Jan;42(Database issue):D1167–75.
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title